X-RAY STRUCTURES OF P22 C2 REPRESSOR-DNA COMPLEXES:

THE MECHANISM OF DIRECT AND INDIRECT READOUT

A Thesis
Presented to
The Academic Faculty

by

Jason Derrick Watkins

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
December 2008

X-RAY STRUCTURES OF P22 C2 REPRESSOR-DNA COMPLEXES: THE MECHANISM OF DIRECT AND INDIRECT READOUT

Approved by:

Dr. Loren Dean Williams, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Stephen Harvey
School of Biology
Georgia Institute of Technology

Dr. Donald Doyle
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Roger Wartell
School of Biology
Georgia Institute of Technology

Dr. Nicholas Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: August 14, 2008

This work is dedicated to my Mother and Father from a grateful son. Thank you.
I love you both.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my advisor and mentor Professor Loren Williams for his invaluable advice and guidance, room to grow as a scientist, and countless hours of work. I would also like to acknowledge Professor Gerald Koudelka for his collaboration and for providing the P22R NTD protein used in this research. Dr. Kris Woods determined the initial crystal growth conditions for the P22R NTD-DNA complex. Discussions and instructions from Dr. Anton Petrov of Professor Steve Harvey's lab were instrumental in the electrostatic surface potential calculations.

I would also like to thank Professor Nick Hud for many helpful discussions on this project, for providing encouragement and advice, and for being an excellent teacher. Professor Allen Orville provided very helpful instructions in x-ray crystallography both here and at Brookhaven National Lab, and challenged me at every turn. I would also like to acknowledge Professor Raquel Lieberman for supporting me while I finished this work as well as in future endeavors. Also, thanks to Dr. Mary Peek for being a great boss and a better friend.

In addition to Kris, I would like to acknowledge the past and present group members who have contributed to my success at Georgia Tech, Tatsuya "T" Maehigashi, Srividya Mohan, Tinoush Moulaei, Chiaolong Hsiao, Seiji Komeda and Shelly Howerton of the Williams' group, and George Lountos of the Orville group.

I would also like to acknowledge Mr. Wayne Findley and Mr. Sonny Kirby for my earliest inspiration and for nurturing a young mind. Your influence is only out

weighed by my appreciation. Thank you both for being, in the truest sense of the word, Teachers.

I would like to acknowledge my Mother and Father, Donna and Bill Watkins, for relentless encouragement, unwavering faith, and unconditional love. Without you both, I would not have persevered.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The P22 c2 repressor protein (P22R) binds to DNA sequence-specifically and helps direct the temperate lambdoid bacteriophage P22 to the lysogenic developmental pathway. To gain insight into its DNA binding mechanism, we solved the 1.6 Å x-ray structure of the N-terminal domain (NTD) of P22R in a complex with a DNA fragment containing the synthetic operator sequence $[d(ATTTAAGATATCTTAAAT)]_2$ This operator has an A-T at position 9L and T-A at position 9R and is termed DNA^{9T}.

Van der Waals interactions between protein and DNA appear to confer sequence-specificity. The structure of the P22R NTD – DNA^{9T} complex suggests that sequence-specificity arises substantially from interaction of a valine with a complementary binding cleft on the major groove surface of DNA^{9T}. The cleft is formed by four methyl groups on sequential base pairs of 5' TTAA 3'. The valine cleft is intrinsic to the DNA sequence and does not arise from protein-induced DNA conformational change. Protein-DNA hydrogen bonding plays a secondary role in specificity.

The structure of the P22R NTD in a complex with DNA^{9C} operator containing the sequence $[d(ATTTAAGACGTCTTAAAT)]_2$ was determined to investigate the effect of a mutation at the central step of the non-contacted region. This operator has a C-G at position 9L and G-C at position 9R. The TpA→CpG mutation at this position is known from solution studies to reduce binding affinity of the repressor. The structure reveals a less stable spine of hydration. Indirect readout appears to be a result of the stabilization or destabilization of the spine of hydration. The difference in stability is related to the sequence in the non-contacted region.

It is known that the non-contacted bases in the center of the complex are important determinants of affinity. The protein induces a transition of the non-contacted region from B-DNA to B'-DNA. The B'-state is characterized by a narrow minor groove a spine of hydration, and localized monovalent cations in the minor groove. The electrostatic surface potential of the P22 c2 repressor DNA binding face is complementary to the narrow minor groove of DNA in the B' state. The free energy of the transition from B- to B'-DNA is known to depend on sequence. Thus the observed DNA conformation and hydration allows formulation of a predictive model of the indirect readout phenomenon.

Heavy atom derivatives of the P22R NTD-DNA$^{9T(C)}$ were crystallized and the cation position was determined from the anomalous and isomorphous difference peaks. Monovalent cations are known to localize in the minor groove with sequence dependence. The localization of cations is linked to the B'-state. However, no peaks from the heavy atom derivatives were observed in the region of the repressor induced B'-state of the operator. The lack of monovalent cation sites observed may indicate i) the presence of a Na^+ specific site in the minor groove, ii) a lack of coordination of cations resulting in the delocalization of cations in the non-contacted region, or iii) the absence of cations in the minor groove of the non-contacted region resulting in an unknown component of indirect readout.

CHAPTER 1

INTRODUCTION

The function of DNA is highly dependent on interactions with proteins. DNA and protein interactions can be both specific and non-specific. Non-specific binding to DNA involves hydrogen bonds and salt bridges between the protein and the DNA phosphate backbone and is independent of the sequence of the DNA. Specific binding is dependent on a proteins ability to discriminate between base pairs to recognize a specific sequence of DNA. The recognition of a duplex DNA sequence by proteins plays a critical role in mediating transcription, recombination, replication and DNA packaging. DNA recognition is accomplished by a combination of two broadly-defined mechanisms: direct readout and indirect readout.

Direct readout of a DNA sequence is defined as the sensing of a base identity by interactions directly with the DNA base. Direct readout is the most recognizable form of DNA-protein interactions, and it typically involves the hydrogen bonding and van der Waals interactions of protein sidechains with the functional groups of DNA base pairs (*1, 2*) and at times may include water mediated interactions.

In indirect readout, DNA affinity depends on the identity of bases not directly contacted by the protein. Indirect readout utilizes relationships between DNA sequence 'deformability' (for example, see *3*). Consequently, indirect readout allows the sensing of DNA sequence at a distance, as first illustrated by the 434 repressor(*4*). The 434 repressor(*4-6*), TATA binding protein (*7-10*), and trp repressor (*11, 12*) are among the first proteins shown to indirectly read (recognize) portions of their DNA binding sites. DNA-binding proteins that are known to use indirect readout for recognition now extend

over the entire tree of life, and include met repressor (*13*), IHF/Hbb (*14, 15*), c-myb (*16*), mar A (*17*), papillomavirus E2 protein (*18, 19*), estrogen receptors (*20*), CAP (*21*), HincII restriction endonuclease (*22*), P22 c2 repressor (*23*) and Ndt80 (*24*).

Indirect readout exploits the complex relationships between DNA sequence, conformation, deformation, hydration, and electrostatics. Despite the prevalence and biological importance of indirect readout, predictive mechanistic models of this phenomenon are not available.

The P22 c2 Repressor

The P22 c2 repressor protein is a member of the lambdoid family of repressors (*25*). This family of repressors plays a fundamental role in determining if a phage will enter a lytic or lysogenic phase of development (*26*). As with other members of the lambdoid family, P22 c2 repressor binds as a dimer to one of six operators that are separated into two regions of a phage plasmid, designated O_R and O_L. Each region contains 3 different operator sites, O_R1, O_R2, O_R3, for the O_R region and O_L1, O_L2, O_L3 for the O_L region (Figure 1.2). The ability of the repressor to discriminate between the three operators of each region is the key to its ability to regulate phage development(*27*).

<u>The P22 c2 Repressor and Indirect Readout</u>

The P22 c2 repressor dimer binds to a pseudosymmetric sequence comprised of 18 bp. The N-terminal DNA binding domain (~ residues 1 to 68) of each monomer contacts the outside seven base pairs and the four central base pairs, two on each half site, between each monomer remaining non-contacted by the protein (*23*). Positions 1, 3, 5, 6, and 7 of the seven bases in the contacted region of each half site of the operators are highly conserved among the different operators, but positions 2 and/or 4 show some

variation between each operator (*23*). While the differences in the non-conserved bases in the contacted region provide some mechanism for the repressor to discriminate between operators, it has also been shown that differences in the non-contacted bases of the operators change the affinity of the repressor (*28-30*). By constructing two operators, DNA^{9T} and DNA^{9C}, it was demonstrated that a TA→CG at the central base step lowered the repressor affinity approximately 10 fold (*28, 29*).

The P22 c2 Repressor and Direct Readout

The P22 c2 repressor is closely related to salmonella phage protein 434 repressor. These repressors share ~30% sequence identity, and the 1985 helix swap experiment by Ptashne showed that both repressors bind through a similar mechanism and architecture(*30*). The similarity between the P22 c2 and 434 repressors was even further confirmed by comparing the NMR structure of the N-terminal domain of each repressor, which superimpose with a rmsd of 2.6 Å (*31*). Both repressors contain five α-helices, with helix 2 and helix 3 forming a helix-turn-helix (HTH) motif (*31, 32*). All lambdoid repressors, as well as many other DNA binding proteins, use the HTH motif for DNA recognition. The x-ray crystal structure of 434/operator complexes show that helix 3 is the recognition helix of this repressor making all base specific interactions with the operator by inserting into the major groove (*32-34*). Biochemical data have indicated that helix 3 of P22 c2 repressor is also responsible for its specificity (*26, 30*).

Extensive mutagenesis experiments by Koudelka of both the P22 c2 repressor and the contacted region of its operator were used to identify which residues of the repressor were responsible for the discrimination of which of the conserved bases (*28, 35*). Using these data and the 434/operator complexes as a guide, Koudelka proposed a model of

P22/operator binding. This model suggests interactions that explain the recognition of the conserved bases of the operator by the repressor (*35*). However, it does not fully demonstrate a reason for the differential affinities of the different operators, nor does it give an indication of the origins of the "indirect readout" mechanism that appears to influence the binding affinity of the repressor.

The following work describes the structure of the P22 c2 repressor (P22R NTD) – DNA complex that reveals a mechanism of indirect readout. DNA is a polyanion capable of taking on multiple conformations(*36*). The propensity of DNA to take on a specific conformation is dependent on sequence specific interactions with water and cations as described below. Indirect readout by P22 c2 repressor occurs as a consequence of interactions of the protein with water molecules and the phosphate backbone of the DNA that induce a B' state. In unbound DNA, the B' state is found exclusively in sequences that contain an A-tract (see below), and is characterized by a narrow minor groove compared to B-form DNA, high propeller twisting, a spine of hydration, and localized monovalent cations in the minor groove (see below and (*37-39*)). An induced spine of hydration in the minor groove of the non-contacted bases appears to be a consequence of solvent stabilization by interactions of water molecules with the protein. The protein also provides an electrostatic surface that is complementary to the B' state DNA. While the protein is capable of enhancing the B'state, the propensity of DNA to adopt the B'state is also dependent on the DNA sequence (see below and (*40-43*). Because of the importance of DNA conformation to indirect readout, the following sections are included to provide the background that details the importance of sequence, water molecules, and solvent cations to the conformational stability of DNA.

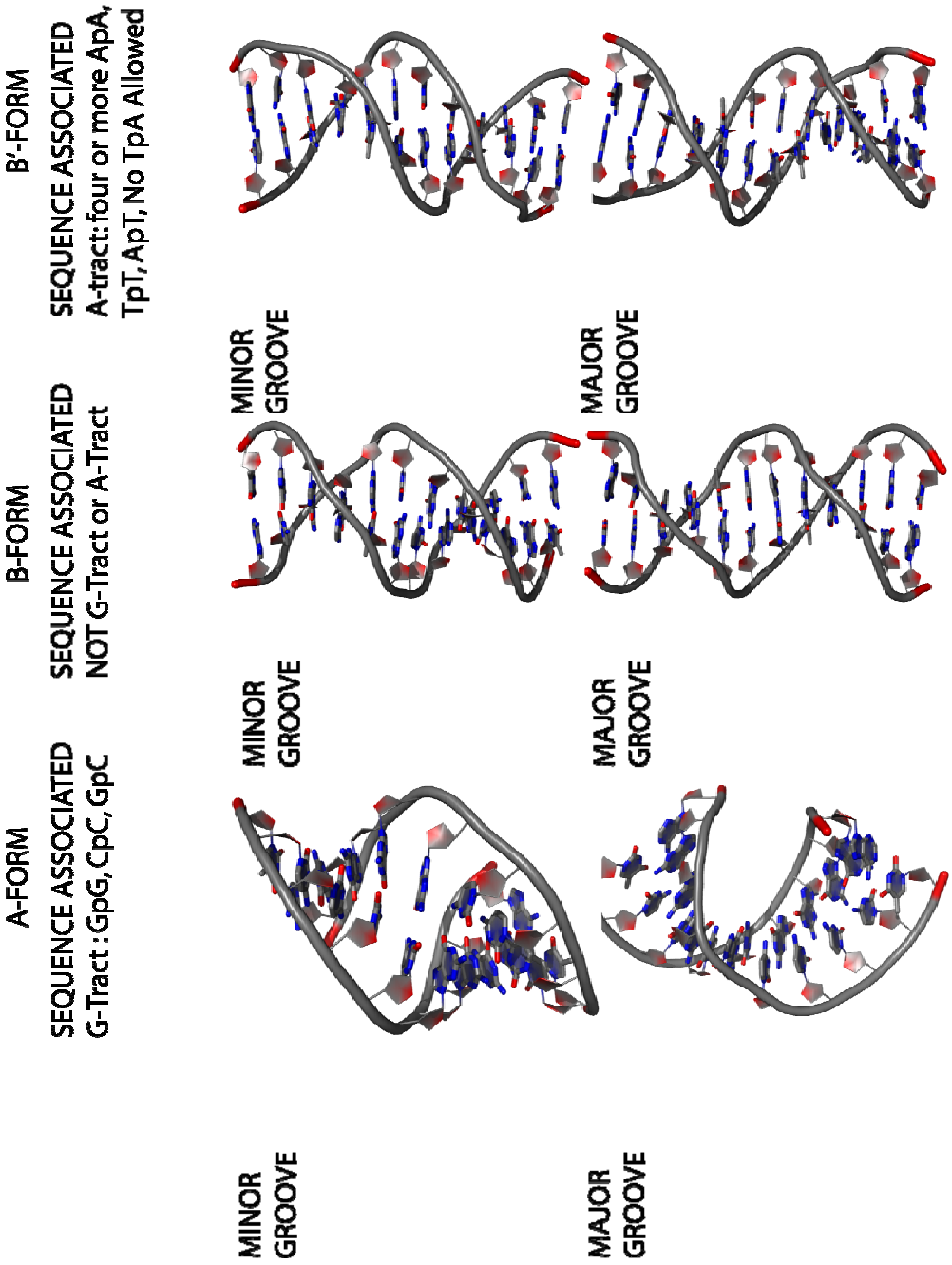

Figure 1.1 Conformations of DNA. A- form pdbID 440D, B-form pdbID 423D, B'form pdbID 355D.

DNA conformation

The classical view of duplex DNA consist of a right handed helical ladder in which the sides are composed of a polymerized deoxy-ribose-phosphate moiety and the rungs of the ladder are the DNA bases (*44*). However, DNA is a dynamic molecule and consists of a variety of conformations including A-form, B-form, and B'-form DNA (reviewed by Hud and Plavec in (*36*)) as well as the left handed Z-form (*45*). The propensity of DNA to obtain a certain conformational state is related to the identity of nucleotide base pairs that make up a specific sequence. Sequences containing G-tracts (repeated GpG, CpC, or GpC content that is rich in GG steps) show a higher propensity to conform to A-form DNA. Sequences containing A-tract DNA (sequence of four or more A's or T's or mixtures of A and T without an TpA step) show a higher propensity to be in the B' state. DNA sequences that contain neither an A-tract nor a G-tract, which is most DNA, favors the B-form conformation (Figure 1.1).

DNA sequence specific cation localization

DNA is a polyanion due to the charged phosphates of the backbone. According to the polyelectrolyte model, DNA is uniformly surrounded by a diffuse cloud of cations (*46, 47*). However, electrostatic surface potential calculations and the pattern of cation localization suggest that the functional groups on the faces of the DNA bases give rise non-uniform electrostatic potential peaks and troughs that are dependent on the sequence (*36, 48, 49*).

Recent work has indicated that groove width is positively correlated with the localization of monovalent cations (*48, 50, 51*). Both x-ray crystallography and NMR have shown that monovalent cations localize in the minor groove of A-tract DNA (*52*-

56). Monovalent cations have also been observed in the major groove at G-tracts (*36, 50, 54*). As discussed below, it appears that two components are necessary for the localization of cations: electrostatics of the groove and the coordinating functional groups present in the groove.

Cation localization in the minor groove of DNA

The minor groove of sequences containing only A•T and/or T•A base pairs, including A-tracts, has exclusively electronegative groups. These sequences give rise to a more negative electrostatic potential in the minor groove. The more negative potential of the minor groove correlates with the position in which cations have been observed by x-ray crystallography and NMR. Conversely, G•C base pairs insert partially positive amino groups into the minor groove. The positive potential in the minor groove of sequences that contain G•C base pairs results in an unfavorable electrostatic environment for cation localization (*36, 49*).

Although all sequences containing only adenine and thymine bases have exclusively electronegative groups in the minor groove, monovalent cations localize only in the minor groove of A-tracts and not in the minor grooves of AT rich non-A-tract sequences. The discrepancy in the localization of cations between A-tract DNA and AT rich non-A-tract DNA cannot be explained by differences in the electrostatic potential. Therefore, it is likely that the difference in the identity and position of metal chelating groups associated with the minor groove also contributes significantly to the localization of cations.

At an ApT step of A-tracts, the O2's of cross-strand thymines are inserted into the minor groove. The lack of a base pair hydrogen bonds and positioning of the O2's of

the thymines at ApT steps create an excellent coordination site for a monovalent cations. X-ray crystallography and NMR have indicated that monovalent cations have 20-50% occupancy at this site (*52-55*, *57*).

Duplex sequences of consecutive ApA or TpT bases would also create a metal coordination site between O2 of thymine and the N3 of a cross-stand adenine. The coordination of a metal ion at ApA or TpT step would likely be less stable than that at an ApT step. The reduced stability of the ApA of TpT site is due to the less electronegative coordinating nitrogen and the close proximity of the C2 of adenine to the coordinated cation. Crystallographic and NMR studies indicate that cations are at a lower occupancy at ApA and TpT coordination sites than at ApT coordination site (*54*).

The introduction of a TpA step would place cross-strand N3's of adenine in the position to coordinate a monovalent cation. While the cross-strand nitrogen would likely be able to coordinate the cation, the site would be the least stable coordination site in an AT rich region. The coordination by two nitrogens and the close proximity of the C2's of the cross-stand adenines may result in a less stable coordination site. Thus, the presence of an O2 from thymine may be necessary for the coordination of cations in the minor groove.

The N2 amino in the minor groove of GC rich regions causes poor coordination of cations in the minor groove. For instance, the coordination of a monovalent cation by the O2's of cross-strand cytosines at a GpC step would be disrupted by the presence of the N2 of guanine. The same is true for sequential CpC or GpG steps in which the coordination of cations would be maintained via the N3 of guanine and the O2 of cytosine. The cross-stand N3's of guanine may be capable of coordinating a cation at a

CpG step similar to that of the N3's of adenine at a TpA step. However, the positive electrostatic potential of the exo-cyclic amines of guanine would make it unlikely that a cation would enter the minor groove to this position. Furthermore, the coordination of a cation by the cross-strand N3's of a TpA step is likely the least stable coordination site in a AT rich sequence. The reduced stability CpG coordination site would be even more substantial than a TpA site when the positive electrostatic potential of the N2 amino group of guanine is considered. For these reasons, it would be unlikely that a cation would localize in the minor groove of GC rich sequences.

Cation localization in the major groove of DNA

The highest occupancy of monovalent cations observed in x-ray crystallography experiments are in the major groove of G-tracts (*50, 54*). The difference in coordination in the major groove arises mainly from the order of hydrogen bond donors and acceptors across a base pair. The order across a G•C base pair would be acceptor-acceptor-donor, and for a C•G base pair would be donor-acceptor-acceptor. This sequence of hydrogen bond donors and acceptors would allow the coordination of a cation between the two sequential acceptor groups. The coordination would be further stabilized by stacking of acceptors at GpG or CpC steps.

The cation coordination site in the major groove created by sequential electronegative groups across G•C or C•G base pair is lost in A•T or T•A base pair. The order of hydrogen bond donors and acceptors for an A•T or T•A base pair is acceptor-donor-acceptor. The placement of donor between the two acceptors likely disrupts any stability in cation coordination.

Unlike the minor groove, the electrostatic potential of the major groove of GC rich DNA is very similar to the potential of AT rich sequences. Both sequences have an equal number of electrostatically equivalent groups. However, the adjacent hydrogen bond acceptors of G•C base pairs may create negative potential "hot spots" at this position in the major groove. These "hot spots" would increase the occupancy of the cations at this position. The negative "hot spots" would be lost in the presence of an A•T base pair. Thus, the coordination and the electrostatics of the major groove favor the localization of cations in G-tracts as observed in x-ray crystallography experiments.

Effects of cation localization on DNA conformation

The localization of cations into a groove would result in an asymmetric decrease in electrostatic repulsion of the phosphate backbone compared to that of a diffuse evenly distributed cation cloud (*48, 49, 58*). The electrostatic collapse model proposes that the phosphate backbone narrows toward the groove that contains localized cations. The difference in the interaction of cations with the different grooves, termed the "tug of war" model by Hud and Plavec, is dependent on sequence (*36*). This explanation is in accordance with the observation that A-tracts have a propensity to obtain the B'state, which is characterized by a more narrow minor groove compared to B-form DNA. Furthermore, this explanation accounts for the propensity of G-tracts DNA to adopt an A-form conformation with a narrow major groove.

Spine of hydration

The minor groove of A-tract DNA has a distinct hydration pattern known as the spine of hydration. The spine of hydration was first observed by Dickerson and coworkers (*37*) and consists of a primary layer and secondary layer of water molecules.

Higher resolution structures also revealed a tertiary and quaternary layer of water molecules (*38*). The primary layer of the spine interacts directly with the functional groups of the bases on the floor of the minor groove. The secondary layer forms hydrogen bonds to the primary layer and to the tertiary layer of the spine. The tertiary layer is capped by the water of the quaternary layer at the lip of the minor groove forming a "fused hexagon" motif that spans the minor groove of the A-tract.

The spine of hydration is associated with the narrow minor groove of the B' state. The spine has been proposed to be a fundamental part of the helix and contribute to the stability of the helix (*40-42*). The spine contributes to the stability of the B' state by protecting the DNA from base-pair opening (*59*). It has also been suggested that the spine of hydration is linked to the localization of monovalent cations (*40*). As discussed above and according to the "tug of war" model, the localization of monovalent cations into the minor groove results in reduced electrostatic repulsion of phosphates across the minor groove, and thus, a narrower minor groove. Thus, there appears to be a thermodynamic cooperation between the localization of cations, the spine of hydration, and the narrow minor groove observed in the B' state.

Solution studies indicate that the stability of the spine of hydration is dependent on the identity and positioning of the functional groups on the floor of the minor groove (*43, 59*). The arrangement of the functional groups of the minor groove required for the primary layer of the spine has the same general hierarchy as that for the localization of cations. The hydration in the minor groove is most stable when coordinated by two cross-strand O2's of thymine. A stable spine is also achieved by the coordination of water between O2 of thymine and a cross-strand N3 of adenine. However, the latter

coordination may result in a lower stability for the bound water than that of the former. Similar to the localization of solvent cations, it appears that the coordination of water by cross-strand N3's of adenine is not stable enough to induce the formation of the spine of hydration in unbound DNA. Thus, a TpA step disrupts the spine of hydration in unbound DNA.

The Spine of Hydration and Indirect Readout

In the work presented here, we propose that the mechanism of indirect readout is dependent on the stability of the spine of hydration. As discussed above, the spine of hydration is an integral part B' state. The DNA operator of the P22 c2 repressor must conform to the B' state for proper orientation with the repressor. In the structures presented in this work, a spine of hydration is observed in the minor groove of the DNA of the A-tracts at each end of the operator and in the non-A-tracts in the center of the operator. The spine of hydration in the center of the operator is stabilized in part by interactions of the tertiary layer of the spine with the repressor. We propose that the stabilization of the spine of hydration by the repressor aids in stabilizing the B' state at the center of the operator. Additionally, the stability of the spine of hydration at the center of the operator is also dependent on the functional groups of the minor groove (see above). Thus, the affinity of the P22 c2 repressor for a DNA operator is modulated by the sequence at the center of the operator despite a lack of direct contacts between the repressor and the bases in the region.

CHAPTER 2

P22 C2 REPRESSOR-DNA^{9T} OPERATOR COMPLEX

To direct the temperate lambdoid bacteriophage P22 to the lysogenic developmental pathway, the P22 c2 repressor (P22R) must recognize and bind six naturally occurring binding sites with different affinities (*60, 61*). The six sites are arranged within the phage into two operator regions, O_R and O_L (*27, 62*). The P22R binding sites are designated O_R1, O_R2, O_R3, in the O_R operator and O_L1, O_L2, O_L3 in the O_L operator. The sequences of the naturally occurring sites (*29*) are partially symmetric . The sequences of the symmetrically-arrayed outermost base pairs in these sites are highly conserved at positions 1,3,5,6 and 7, while positions 2 and 4 are variable (*23, 62*). Furthermore, The central four bases of the operator also show a high degree of divergence.

Biochemical experiments suggested that conserved bases in the naturally occurring sites (*29*) are contacted by bound protein (*29, 35*). Similar experiments indicate that the variable central bases are not contacted by P22R. Yet, changes in the sequence of these non-contacted bases alters the operator's affinity for protein (*23, 28, 29, 35*). For example T-A→C-G transitions of operator 9T at positions 9L and 9R (Figure 2.1) decrease P22 c2 repressor affinity by 10-fold (*23*). Hence, P22R protein recognizes its specific DNA sites by directly contacting the bases in the outer regions of the operator and discriminates between sites through indirect readout, by sensing the sequences of the central non-contacted regions.

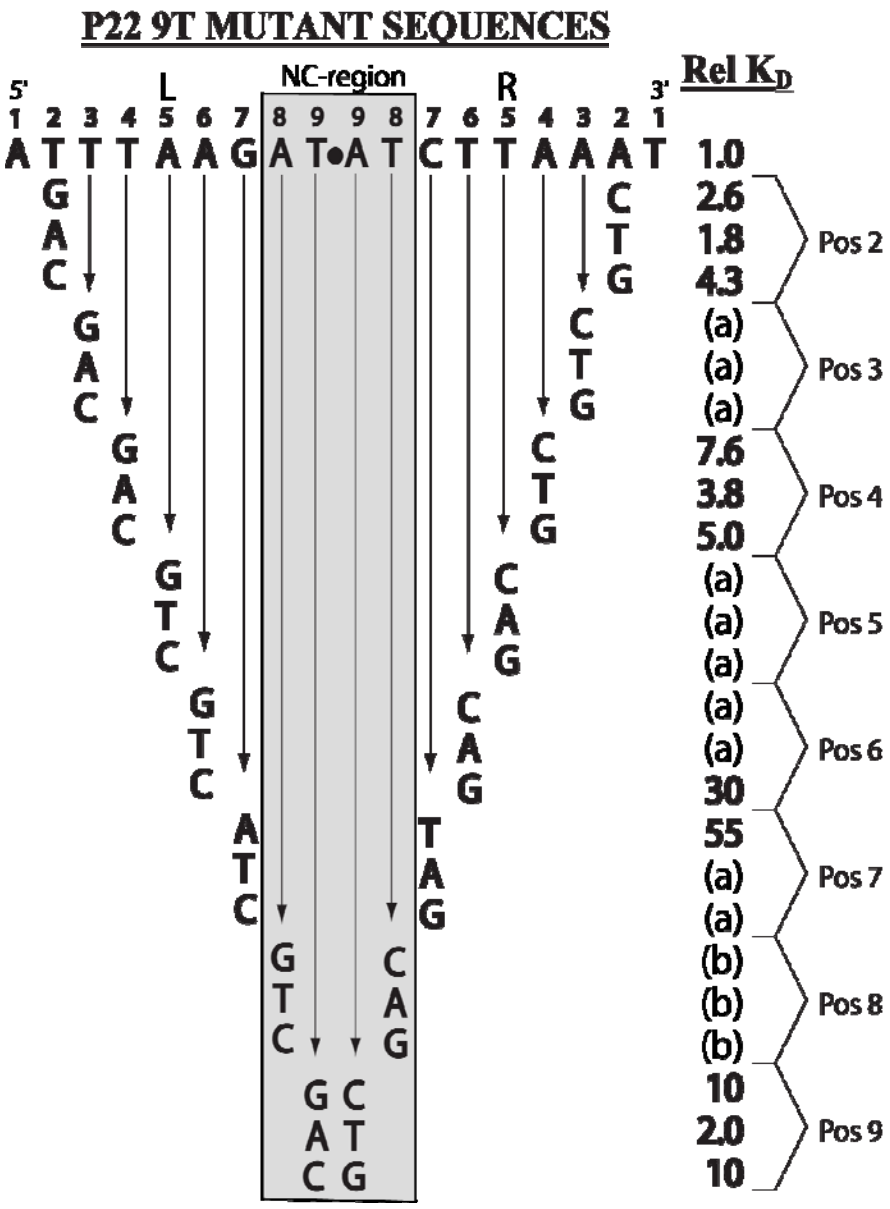

Figure 2.1. The consensus P22R binding sequence (synthetic) and mutants of the synthetic binding site. Binding affinities of P22R are expressed relative to that of the consensus synthetic operator. The Rel K_D is the relative dissociation constant at 100 mM KCl normalized to the tightest binding synthetic operator, 9T. The apparent K_d (K_D) is defined as the concentration of repressor monomers needed to half-maximally occupy a binding site. Under the conditions of these experiments, P22 repressor would be anticipated to be 100% monomeric. hence, any differences in apparent K_D reflect changes in the stability of the repressor-DNA complexes. The value used for normalization is 1= 1.6×10^{-8} M. The gray bar highlights the positions of the noncontacted base pairs. (a) indicates that the repressor binds to the operator only nonspecifically. Nonspecific binding is observed at concentrations $< 2 \times 10^{-6}$ M. (b) indicates data are not available.

To decipher mechanisms of both direct and indirect readout, we solved the x-ray structure of the P22R N-terminal DNA binding domain (P22R NTD, residues 1 to 68) in complex with the synthetic 9T operator (DNA^{9T}). DNA^{9T} has an T-A base pair at position 9L and A-T at position 9R (see Figure 2.1).The structure of the complex allows us to extend and refine a binding model of Koudelka (*35*). This model is based on extensive mutagenesis and biochemical data, the NMR structure of the P22R alone (*31*) and homology modeling, using the 3D structures of 434 repressor-operator complexes (*32-34*) as templates. The model predicts specific contacts and attempts to account for highly conserved and variable positions of the operator. The three dimensional structure described here allows us to test and extend the Koudelka model.

Here we describe a novel mechanism of direct readout that utilizes shape complementarity between the major groove of the DNA and the protein 'reading head'. A sequence-specific cleft, formed by the methyl groups of thymine groups on the major groove surface, matches the shape of the recognition helix. Simple non-directional van der Waals interactions between protein and DNA appear to confer sequence-specificity. The hydration and conformation of portions of this structure are consistent with a protein-induced transition from B-DNA to B'-DNA at the central non-contacted region. The free energy of this change in DNA state is known to vary with sequence (*63*). The structure provides a molecular level explanation for indirect readout: a protein-induced alteration in the non-contacted region of the DNA that is supported by interactions between minor groove functional groups of DNA, and water molecules and presumably cations.

MATERIALS AND METHODS

Construction of plasmid expressing P22R NTD

A DNA fragment encoding the NTD fragment of P22 repressor was created using PCR to insert a stop codon at position 69 of the intact P22 repressor gene. We amplified a 450 bp fragment from pTP125 (*64*) using primers obtained from the CAMBI Nucleic Acid Facility (University at Buffalo). The sequences of the primers were: 5'GTTTTTTGCGTCGACATCATA 3' and 5' CGTTTGTCTGAC GTCAATCTCCTTTC 3'. This fragment, containing the tac promoter (*65*) fused to the P22 repressor gene fragment, was cleaved with Aat II and Sal I, gel purified, ligated in pUC18 cut with the same enzymes and transformed into E. coli strain XA90 (*66*) . The resulting plasmid expresses a protein comprised of the first 68 amino acids of P22 repressor.

Purification of P22R NTD

E. coli XA90 cells (*66*) bearing P22R NTD expression plasmid were grown with aeration at 37^{o}C in 3 liters of Luria broth supplemented with 0.1 mg/mL ampicillin to an A_{600} of approximately 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mM, and growth at 37^{o}C was continued for 4 hours. The cells were harvested by centrifugation and resuspended in 20 mL lysis buffer (200 mM NaCl, 50 mM EDTA; 100 mM Tris pH 7.4) with 0.5 mL of 20mg/mL phenyl methyl sulfonyl fluoride (PMSF). The cells were lysed in a French pressure cell (18,000 lb/in^2) at 4^{o}C. The lysate was diluted with 100 mL of cold lysis buffer. Cellular debris was precipitated by centrifugation at 4000 x g. Nucleic acids were precipitated by dropwise addition of polyethyleneimine to a final concentration of 0.6% at 4^{o}C and mixed for 15 minutes, followed by centrifugation

at 10,000 x g. Soluble proteins were precipitated with 40% (w/v) ammonium sulfate and dialyzed at 4^{o}C against three 1 L changes of standard buffer (SPB-50 mM $NaPO_4$ pH 6.8 10 mM EDTA) supplemented with 50 mM NaCl. The dialyzate was loaded onto three Econo-Pac carboxymethyl (CM) cartridges (Bio-Rad, Hercules, CA) and eluted with a linear salt gradient from 0.1 to 1 M NaCl in SPB over 70 minutes. Fractions containing repressor were pooled and precipitated with 40% ammonium sulfate. Following precipitation, the protein-containing pellet was dissolved in SPB+100 mM NaCl and loaded onto a 1 cm X 50 cm Sephacryl S100 (Amersham Biosciences) size exclusion column. The column was developed with the same buffer at a flow rate of 0.3 mL/min. Fractions containing P22R NTD were pooled and the protein was precipitated with 40% ammonium sulfate. The dissolved pellet was dialyzed against SPB supplemented with 150 mM NaCl and 20% glycerol. The protein was flash frozen in liquid nitrogen and stored at -70°C.

DNA purification

DNA oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). The oligonucleotides were purified with denaturing polyacrylamide gel containing 7M urea and TBE (89 mM Tris borate, pH 8.4, 1 mM EDTA) and subsequently dissolved in TEN (10 mM Tris, pH 8.0 50 mM NaCl, 1 mM EDTA). To form the double-stranded protein binding site, equimolar amounts of the complementary strands were mixed, heated to 85°C for 60 s and slow-cooled over four hours to anneal. Double-stranded DNA was separated from the single-strands by electrophoresis on 12% polyacrylamide gels in 1X TBE.

Co-crystallization of P22R NTD with DNA^{9T}

Crystals were grown at 4°C by the hanging drop vapor diffusion method. The initial crystallization solution contained 0.42 mM P22R NTD, 0.42 mM duplex d($^{5'}$TATTTAAGATATCTTAAATG$^{3'}$) -d($^{5'}$CATTTAAGATATCTTAAATA$^{3'}$), 45 mM Tris·HCl (pH 7.8), 19 mM NaCl, 1.9 mM glycerol, 11% PEG 400, 4.5 mM LiCl, 2.3mM $MgCl_2$ and 0.91% MPD in a volume of 5.3 µl. The crystallization solution was equilibrated against a reservoir of 100 mM Tris·HCl (pH 7.8), 25% PEG 400, 10 mM LiCl, 5 mM $MgCl_2$ and 2% MPD. Crystals grew to a size of approximately 0.2x0.2x0.2 mm^3 within one day. A single crystal was collected from the crystallization solution in a nylon loop and flash frozen in liquid nitrogen.

Data Collection

Diffraction data were collected on SER-CAT beamline 22 ID at the Advanced Photon Source, Argonne National Laboratory with a MAR300 CCD detector using radiation of 0.99997 Å wavelength and a crystal-to detector distance of 170 mm. Images were collected with a 1 s exposure time and a 1° oscillation. The crystal was maintained at 110 °K. A total of 347,031 reflections were indexed and integrated with DENZO (*67*) and reduced to 55,132 unique reflections in space group $P4_3$ in the resolution range of 35-1.53 Å. Data were scaled with the program SCALEPACK(*68*).

Refinement

Initial phases were determined by molecular replacement with the program CNS (*69*) using data from 15 Å to 3Å, and the 434 repressor/O_R2 complex (PDB entry 1RPE) as a search model. All non-identical residues (P22R versus 434 repressor) in the search model were mutated to alanine. The 434 repressor NTD shares ~30% sequence identity

Table 2.1. Data Collection and Refinement Statistics

X-ray source	SER-CAT beamline 22-ID APS
Wavelength (Å)	0.99997
Detector	MAR CCD 300
Resolution Range (Å)	35.00-1.53
Mosaic spread (°)	~0.7
Space Group	$P4_3$
Unit-cell parameters (Å)	
a	64.059
b	64.059
c	101.604
Total Reflections	347,031
Unique Reflections	55,132
Multiplicity	5.8 (4.7)*
Completeness (%)	99.4 (99.3)
Highest Resolution Shell (Å)	1.57-1.53
$I/\sigma(I)$	47.1 (2.12)
Rsym (%)	6.9 (65.9)
Reflections (working/test set)	53,765/6,068
Number of non-H protein atoms	2 x 514
Number of non-H DNA atoms	814
Number of water atoms	330
R-crystal	20.35%
R-free	22.53%
Average B-factor	
Protein	31.55 (L) 25.69 (R)
DNA	29.43
Water	42.62
RMSD from ideal	
Bond length (Å)	0.0099
Bond angle (°)	1.33

* The number in parenthesis indicates the value for the highest resolution shell.

with the P22R NTD. Following a rigid body refinement in CNS, numerous cycles of model building based on $|2F_o-F_c|$ and $|F_o-F_c|$ maps were performed in the program O (*70*), followed by refinement in CNS using the R-factor and R_{free} as guides following each cycle. To reduce model bias, simulated annealing was performed during each stage of refinement with a starting temperature of 5000 °C and 25 °C steps in the early rounds of refinement and 2000 °C starting temperature and 50 °C steps in later stages. Water molecules were added iteratively to the model, followed by cycles of refinement. The final refined model of the P22R NTD (residues 1-68) in a complex with DNA^{9T} is refined to 1.53 Å resolution, with a R-factor 20.35 % and a R_{free} of 22.53 %. Data collection and refinement statistics are given in Table 2.1.

Electron Density and Model Quality

The backbone atoms of the protein, as well as the sidechain atoms, are well-ordered and clearly observable in the $2F_o-F_c$ Fourier maps with only a few exceptions occurring for long surface sidechains and the terminal residues. The N-terminal residues (1-2) are apparently disordered and their density is not observed. These four residues (2 per monomer) are not omitted from the model. Ill-defined density is observed for residue 3 and C-terminal residue 68. These residues are built in at half occupancy, which gives the best fit to the data, using the R-factor and R_{free}, as well as sum and difference density maps as guides.

The electron density of the DNA is sharp and well-defined throughout with exceptions at the terminal base pairs (below). The R half site has better-defined density, lower thermal factors and more observable solvent molecules than the L half site. The conformation of the DNA oligomer is not symmetric in that one of the terminal base pairs

is a G·C base pair and the other is an A·T. The R-factor and R_{free} as well as the difference density indicate a best fit with the G·C base pair in the R half site terminus and the A·T in the L half site terminus. The terminal base pairs are probably characterized by multiple conformations and mixed base pair occupancy. The electron density is not well defined at either terminus. The thermal factors of the L half site terminus base pair are highest.Attempts to model multiple conformations were not justified by fitting statistics. Therefore, only single, apparently predominant occupied position and base IDs were modeled for the terminal base pairs.

RESULTS

The Global Complex

The electron density maps of the complex are clean, continuous and detailed (Figure 2.2). P22R NTD binds to DNA^{9T} as a dimer (Figure 2.3). The two subunits of the dimer are related by an approximate 2-fold rotational axis that is roughly perpendicular to the DNA helical axis, bisecting the 18 base pair recognition sequence. Protein interactions with the two DNA half sites are symmetric about pseudo-twofold in the center of operator. With a few exceptions, this symmetry extends down to the level of solvent locations and interactions. Hence the description here is given for the structure and interactions of the R half site (Figure 2.4), but applies equally well to the L half site. The observed symmetry argues against the influence of specific lattice forces on conformation or interactions.

Each protein monomer contains five α-helices made up of residues 6 to 17 (helix 1), 21 to 28 (helix 2), 32 to 40 (helix 3), 47 to 57 (helix 4), and 61 to 65 (helix 5). The x-ray structure of P22R NTD – DNA^{9T} complex is consistent with the Wüthrich NMR

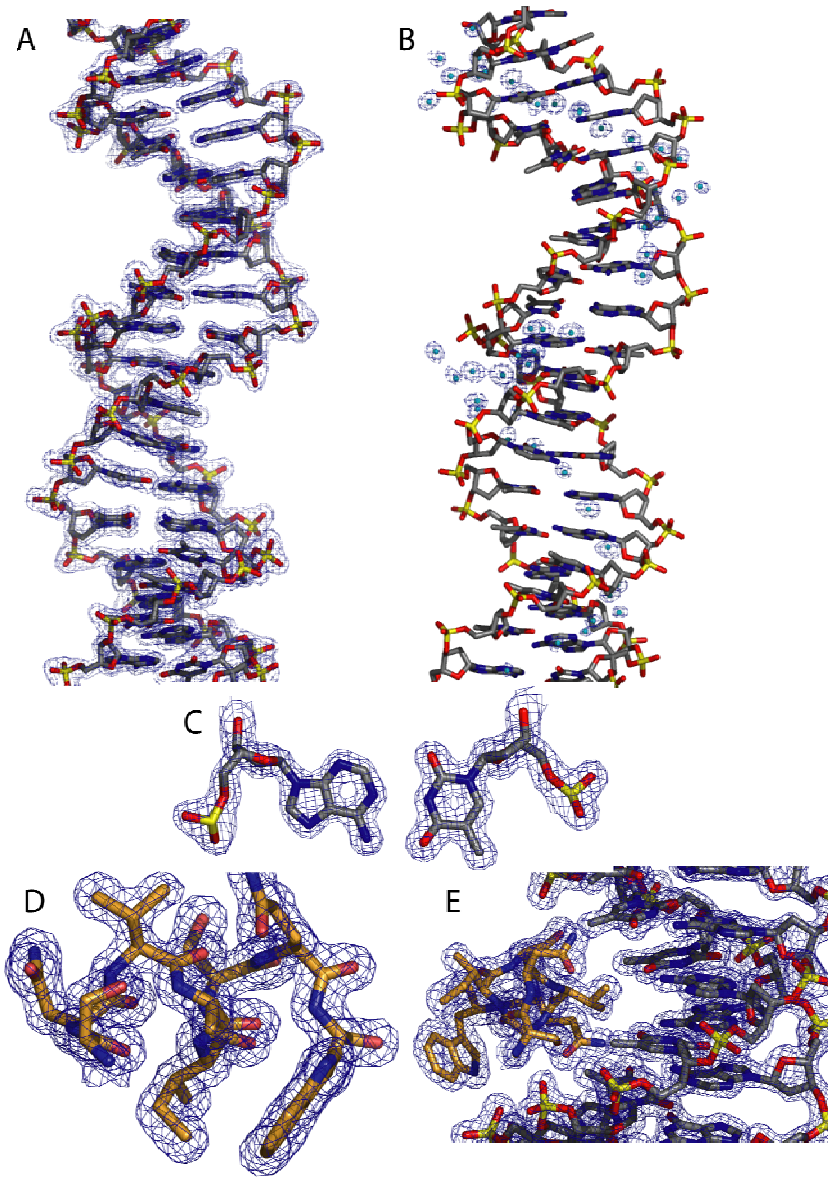

Figure 2.2. The electron density defining the DNA, protein and solvent is of high quality. Shown here is (A) electron density surrounding the DNA duplex contoured at 1.5 σ, (B) electron density surrounding minor groove solvent contoured at 1.3 σ, (C) electron density surrounding base pair T(9)-A(32) contoured at 1.3 σ, (F) electron density surrounding amino acids of the recognition helix (residues 31-40) contoured at 1.5 σ, (E) electron density surrounding the DNA and the R-recognition helix contoured at 1.5 σ. The maps are annealed/refined 2Fo-Fc composite omit maps.

structure of the protein alone (*31*). The protein forms a globular structure with hydrophobic residues in the core and charged and polar residues on the exterior, exposed to solvent and to DNA. As anticipated (*30, 31, 35*) P22R NTD contains a helix-turn-helix (HTH) motif made up of helices 2 and 3. The N-terminus of helix 3, the DNA recognition helix, is further from the central non-contacted region than the C-terminus. Helix 3 is the most ordered region of the polypeptide as indicated by thermal factors and electron density maps. Helices of the N and C termini are more disordered, with solvent exposed residues showing the most disorder. A Ramanchandran plot (*71*) shows that all residues fall in the allowed regions for backbone conformation.

Global DNA Conformation

Perturbations in base pair roll and propeller twist induce complementarities of the global shape of the DNA with the shape of the protein. An overall DNA curvature of 16° allows correct alignment of successive major groove faces with recognition helices (Figure 2.3). The overall conformation of DNA^{9T} is B-form, with directions and positions of axial deviations from linearity allowing the DNA to successfully transit from one recognition helix to the next. Deviations from linearity of the axis are associated with positive rolling between base pairs 5 and 6 (9°) and between base pairs 6 and 7 (11°).

The Non-contacted Region

<u>The Induced B' State</u>

The non-contacted region of DNA^{9T} in the P22R-NDT complex is in the B' state as indicated by a spine of hydration and an especially narrow minor groove (Figure 2.5). The B' state is not normally observed in non-A-tract DNA sequences (ATAT is not an A-

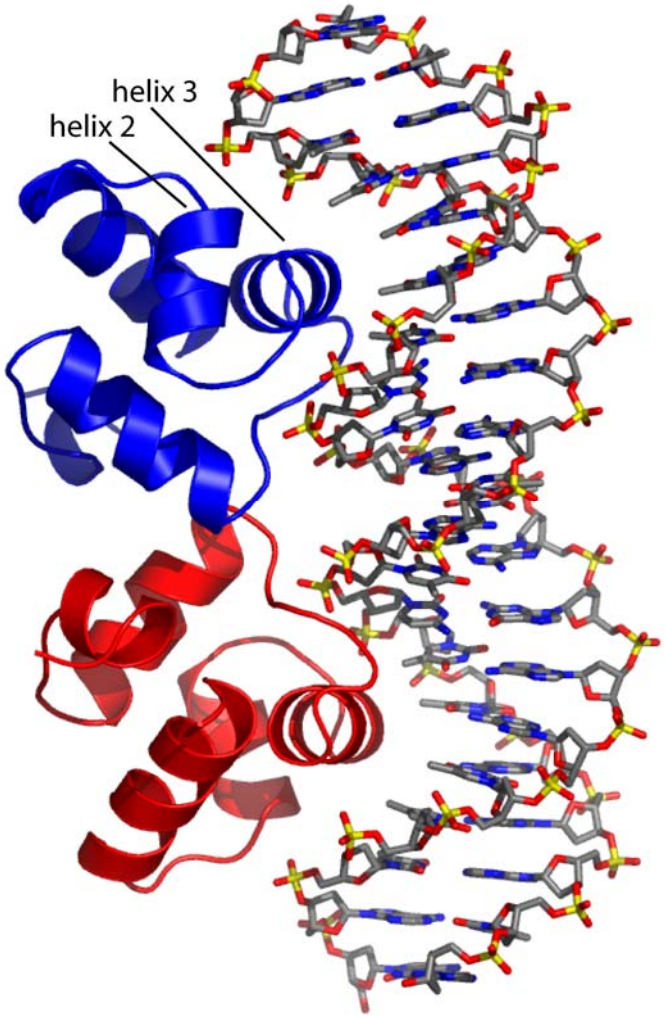

Figure 2.3. (A) Global view of the DNA protein complex. The R-subunit is represented by red ribbon, The L-subunit is represented by blue ribbon. The DNA represented as stick with C, gray; N, blue; O, red; P, yellow.

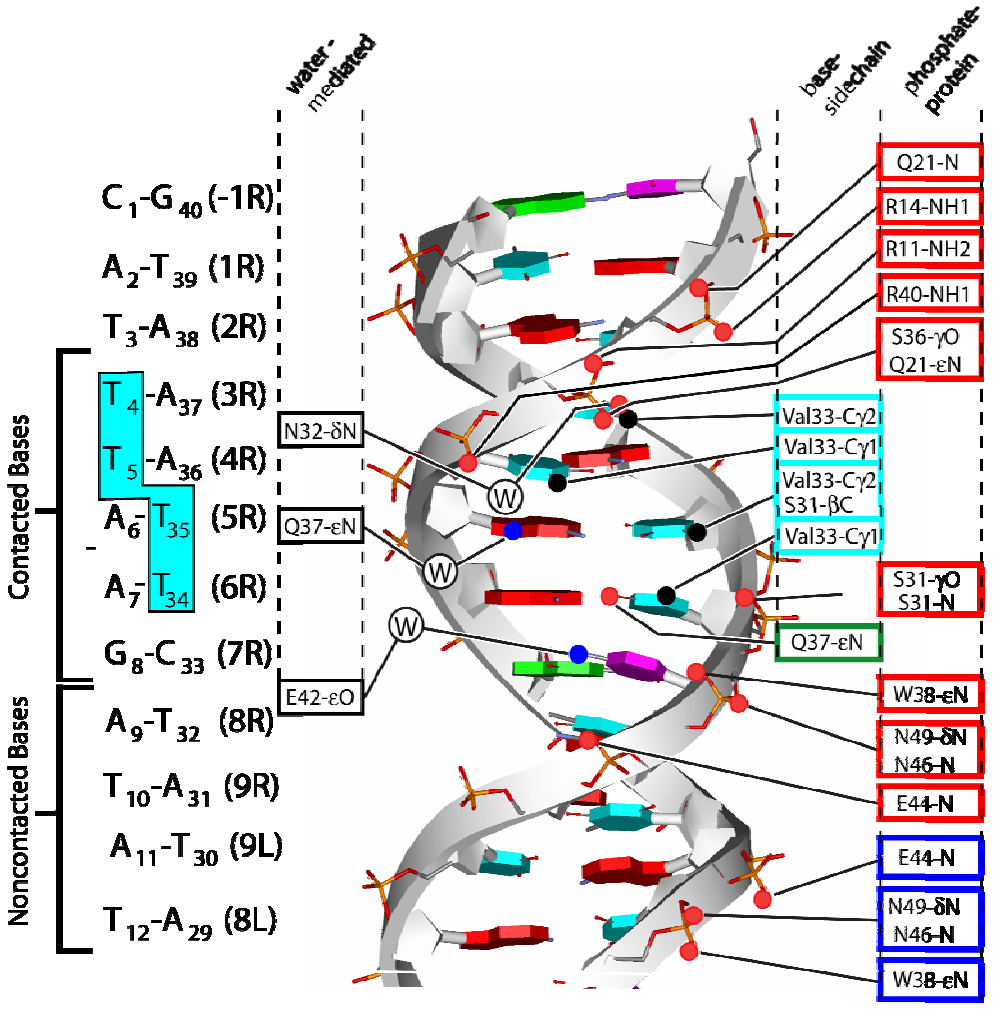

Figure 2.4. Schematic diagram of P22R NTD - DNA^{9T} interactions. Operator positions are numbered (1R, 2R, etc) and labeled as in the PDB file (in subscript). The thymine residues that form the Valine Cleft are highlighted in cyan. Protein atoms that interact with DNA^{9T} are listed in rectangles. The left-hand column gives water-mediated contacts, the central column gives base-specific contacts and the right-hand column gives protein-phosphate contacts. The DNA bases T(cyan), A(red), G(green), and C(purple) are shown in cartoon. The backbone is highlighted by a gray band. tract). However, induction of the B' state in ATAT sequences can be caused by crystal lattice forces (*72*). It appears that the B' state of the non-contacted ATAT sequence of DNA^{9T} is caused here by protein binding. For comparison, Figure 2.6 shows a superimposition of the ATAT elements in the unperturbed B-conformation, a lattice-induced B' state, and our protein induced B' state. The minor groove is narrow in both induced B' state fragments and is wider in the unperturbed state.

Minor groove hydration

Continuous spines of hydration are observed in both the contacted A-tract region of DNA^{9T} comprised of positions R1 through R4 and in the non-contacted region (positions 8L-9L, 8R-9R), which does not contain an A-tract sequence. The base of the spine is a series of water molecules that forms bridges between hydrogen bond acceptors of DNA bases (Figure 2.5A and 2.5B). A secondary layer forms bridges between adjacent waters in the primary layer. At the central TpA step, water 5 bridges two N3 atoms [the N3 of A(11) and the N3 of A(31)]. This N3-W5-N3 bridge at the TpA step, is not a component of A-tract hydration (TpA steps are not allowed in A-tracts). However the N3-W5-N3 interaction very closely mimics the bridging interactions of an A-tract. Like water molecules at ApT steps of A-tracts (*58*), W5 is six-coordinate, interacting with two cross-strand hydrogen bond acceptors on the floor of the groove, two O4' atoms of the deoxyribose backbone and two water molecules in the secondary layer of the spine (Figure 2.5B).

The secondary hydration layer of the non-contacted region minor groove is orderly and well-formed, and supports the characteristic third layer of hydration. The hydration of the L-end of the operator is similar but appears less extensive than that of the R-end. This difference is likely due to the higher overall disorder of L-half-site, which is indicated in the electron density and the higher overall thermal factors.

Counterions

We assume there are counterions within the non-contacted minor groove. However, the differentiation of cations such as Na^+ from water molecules by x-ray diffraction is problematic because of mixed and partial occupancy, similar scattering

properties of Na^+ with water, and positional disorder. All solvent peaks are assigned as water molecules, although it is likely that some positions are at least partially occupied by cations (*54, 73, 74*). Involvement of cations in indirect readout of non-contacted bases of P22R is indicated by biochemical studies (*23, 28*).

Narrow Minor Groove in the Non-Contacted Region

The global minimum in minor groove width is at the center of the non-contacted region (positions L9 and R9, Figure 2.5C). The minor groove is wide from R8 through R5 then narrows from R4 through R1 (the A-tract).

The Tunnel

The DNA^{9T} minor groove and the repressor combine to form a tunnel (Figure 2.7). The floor of the tunnel is the minor groove edges of the bases of residues 11-13, and 31-32. The walls are the sugar-phosphate backbone. The ceiling of the tunnel is the dimer interface, i.e., the loop region connecting helices 3 and 4 and the N-terminal region of helix 4. The tunnel is filled with solvent molecules. The floors, walls and ceiling appear to be impermeable to water and ions. Hence solvent molecules would enter and exit the tunnel only by traversing along the minor groove.

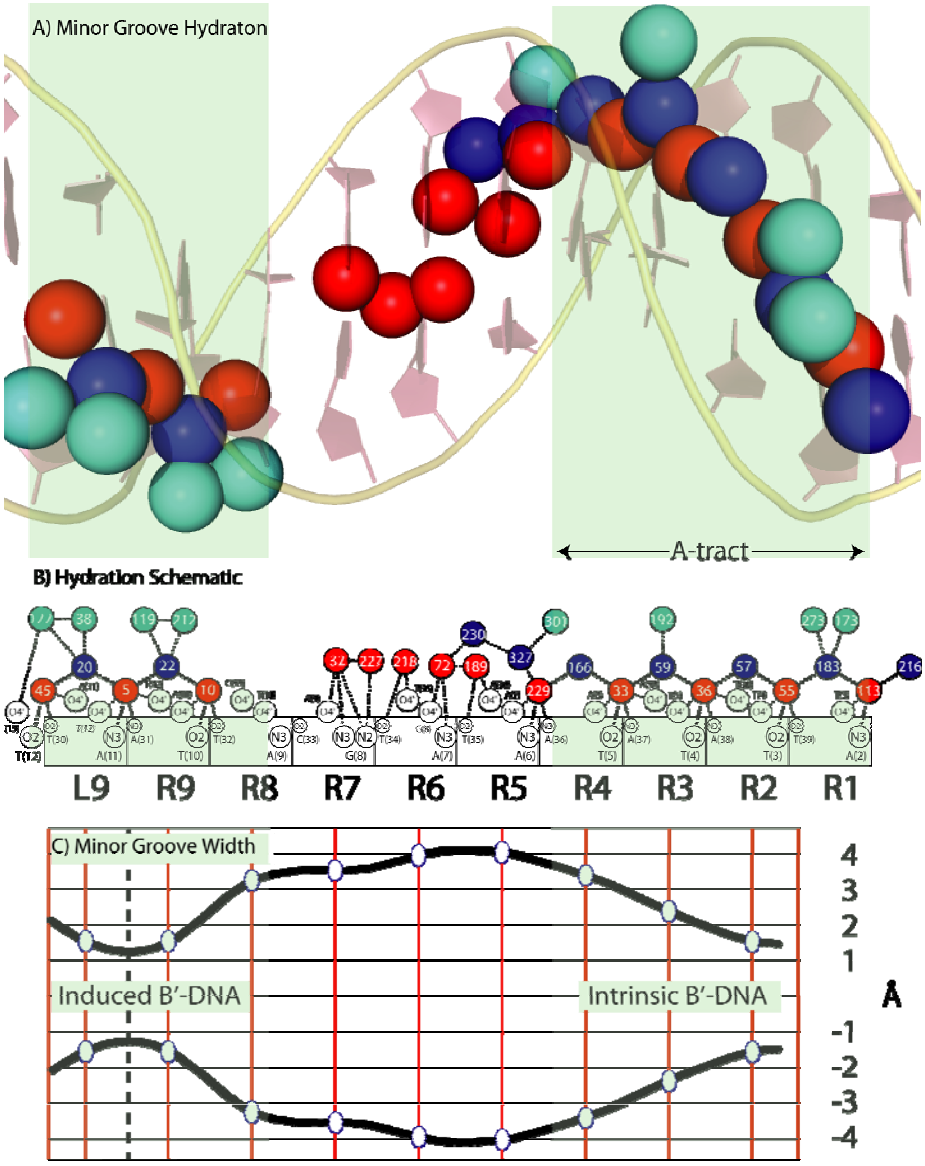

Figure 2.5. The B' state, induced and intrinsic. The R half site of the P22R NTD – DNA^{9T} complex, the A, B & C panels are aligned, with a green background indicating the regions of induced and intrinsic B' DNA. The A-tract is indicated. (A) Minor groove hydration. The primary hydration layer, composed of water molecules in contact with DNA bases, is red. The secondary hydration layer, composed of water molecules in contact with the primary layer, is blue. The tertiary hydration layer, composed of water molecules in contact with the secondary layer, is green. (B) A schematic representation of the minor groove hydration, showing the zig-zag spine of hydration in the induced (non-contacted region) and intrinsic B' regions. The color scheme is the same as in panel A. Hydrogen bonds are indicated by dashed lines. (C) Minor groove width profile, showing the narrow minor groove of the induced and intrinsic B' regions. Each line represents one strand, so that the distance from line to line is the width of the minor groove.

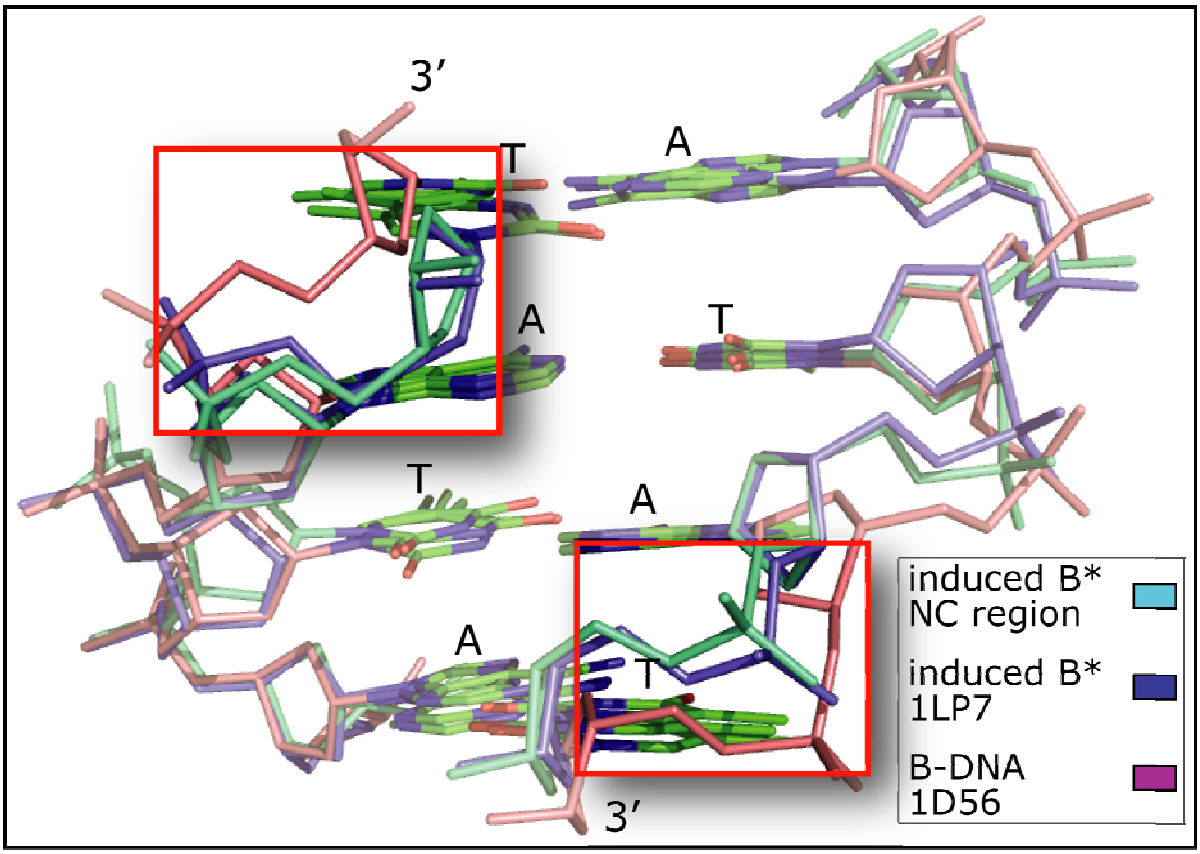

Figure 2.6. A superimposition of ATAT segments in two DNA fragments in which the B' states has been induced, along with DNA in the unperturbed B state. Protein-induced B' state (P22R NTD –DNA^{9T}) is cyan, lattice-induced B' state is blue (PDB code: 1LP7), and unperturbed B state is violet (PDB code: 1D56). The red box highlights the narrow minor groove of the two B' DNA fragments, in comparison with the unperturbed B-DNA.

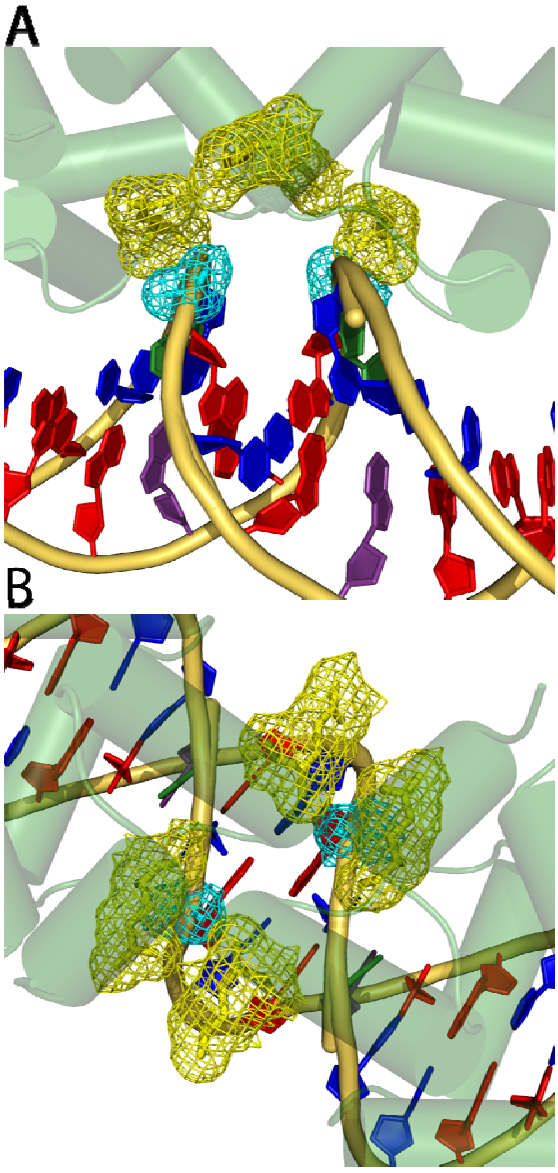

Figure 2.7. The Minor Groove Tunnel. The tunnel is made up of the minor groove of the DNA and dimer interface of the protein. (A) A view down the tunnel, along the minor groove. (B) A view through the protein, across the minor groove. The protein is green with the cylinders representing the α-helices. The DNA is colored red (A), blue (T), purple (G), and green (C). The sugar-phosphate backbone is a brown tube. Glutamates 44 and 48 (yellow) and the proximal phosphate groups (cyan), shown in mesh, emphasize the high electronegative potential of the tunnel.

DNA Base – Protein Contacts

The protein sidechain to DNA base interactions (Figure 2.4) involve Helix 3, as well as several residues on either side of Helix 3. These base-sidechain interactions involve the outer seven base pairs of the operator half site. The two recognition helices, one from each monomer, contact consecutive major grooves, on a common helical face. The protein sidechains of residues 31,32, 33, 37 and 42 contact the functional groups of bases in the contacted region of the operator. Each of these protein residues contacts a base of the operator either directly, as with 31, 33 and 37 or via a water meditated contact as with 32 and 42. Contacts made by these amino acids, which appear to be base-specific, are supported by buttressing interactions with other residues in the recognition helix and by Ser 31, directly preceding the recognition helix. A complex network of supporting and solvent-mediated interactions helps explain the pleiotropic effects that changing residues in the recognition helix has on DNA binding specificity (*28, 29, 35*).

The Valine Cleft

Valine 33, the second residue of the recognition helix, is inserted into a prominent cleft , which we call the Valine Cleft, in the major groove of the DNA (Figure 8). The shape of the cleft is complementary to that of the valine sidechain. Interactions of Val33 within this pocket appear to be important for positioning the protein relative to the DNA, and may confer thermodynamic stability and specificity.

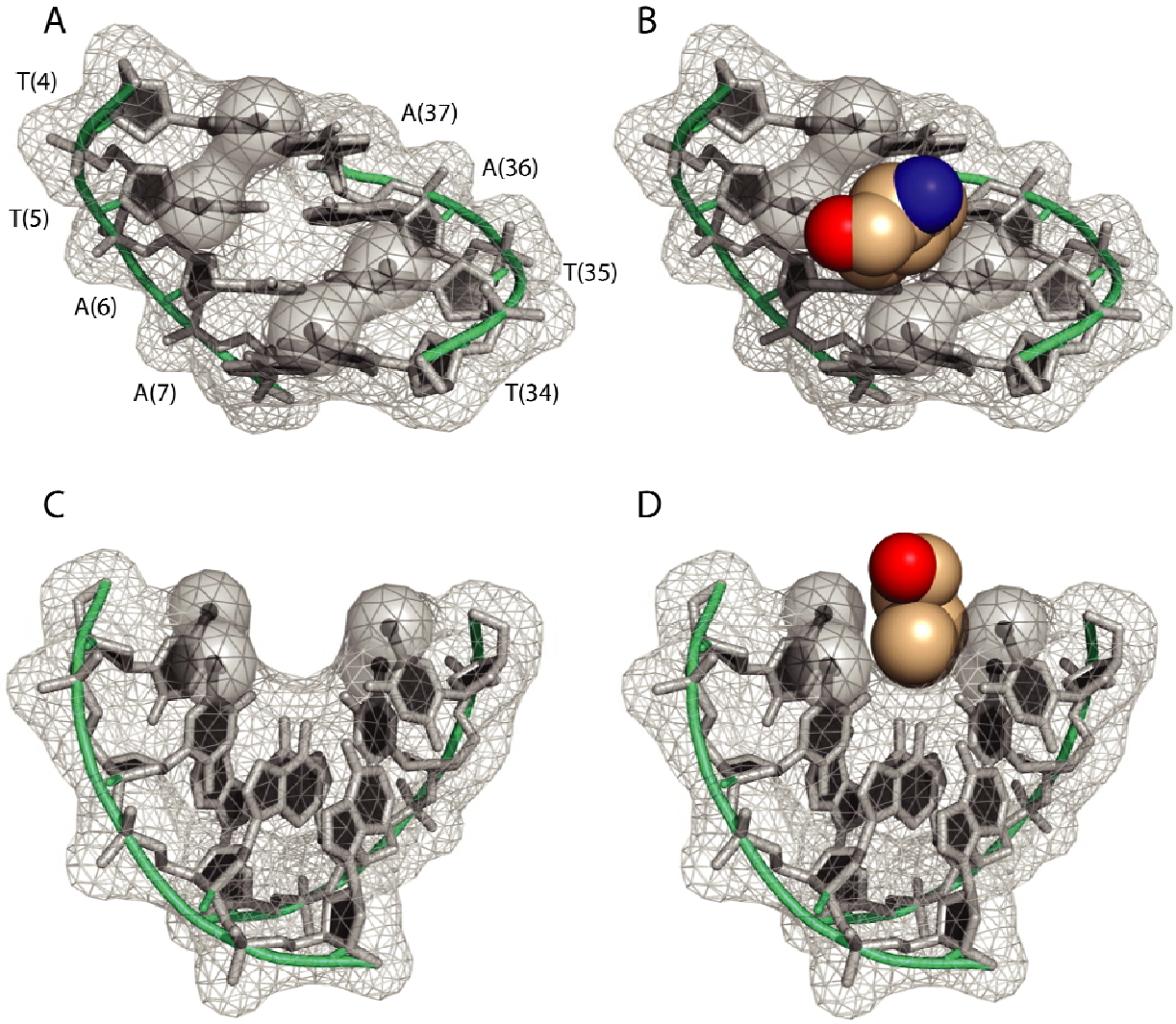

Figure 2.8. The Valine Cleft. The surface of the DNA atoms composing the R-half site Valine Cleft is represented in net, the atoms are in stick representation. A green tube highlights the backbone. The four methyl groups that form the Valine Cleft are shaded. Val 33 is shown in space filling representation, with C, tan; N, blue; O, red. (A) DNA only, viewing into the major groove, showing the 'empty' Valine Cleft. (B) The DNA plus Val 33, showing the shape complementarity of the Val 33 and the Valine Cleft. (C) Same representation as panel A except viewing along the major groove. (D) Same representation as panel B except viewing along the major groove.

The Valine Cleft is composed of two opposing sets of major groove methyl groups contributed by A-T base pairs of the sequence 5' TTAA 3'. The 5-methyl groups of positions 3 and 4 form one hydrophobic wall with contributions from the methyl group of the T-A at position 2. The 5-methyl groups of positions 5 and 6 form a second hydrophobic wall. The cleft appears to be a natural consequence of the TTAA sequence element and is not induced by protein binding. An empty cleft with similar dimensions is observable in an unbound DNA fragment with the same 5' TTAA 3' sequence element (NDB entry BD0051 (*72*)). In DNA^{9T}, negative propeller twists (-11 to -16°) of positions 2 through 7 (except position 5 with a propeller twist of -3°) facilitate collapse of methyl groups on opposing sides of the major groove toward a common center, decreasing their relative offset along the helical axis. Similar propeller twisting is observed in the free cleft (BD0051). The natural propeller twist and base pair roll associated with the AATT sequence are important in forming the cleft in both the free and bound DNA fragments. The floor of the cleft is hydrophilic, and is formed by the O2(T) and N6(A) atoms of base pairs 4 and 5. While desolvation of the hydrophilic floor of the cleft is expected to be thermodynamically unfavorable, the desolvation of the hydrophobic Val33 sidechain and the walls of the cleft would be favorable.

On one end of the cleft, Val33 is buttressed by Asn32 and Ser31, near the N-terminus of the recognition helix. The Oγ of Ser31 donates a hydrogen bond to the O1P of thymine at position 6 (2.7 Å) and accepts a hydrogen bond from NH of Ala34 (3.07 Å). The Cβ and Oγ of Ser31 form van der Waals contacts with the 5-methyl of thymine at position 5 (3.65 Å). This interaction network appears to help stabilize the position of the protein relative to the DNA, helping to fix Val33 within the cleft.

On the end of the cleft closer to the center of the operator, Val33 is buttressed by Gln37 and Ser36. In the single direct hydrogen bond (per half site) between the protein and a DNA base, the ε-NH_2 of Gln37 forms a hydrogen bond with the O4 of thymine at position 6 (3.06 Å). Several water-mediated hydrogen bonds link ε-NH_2 of Gln37 to the DNA, as described below. The sidechain of Gln37 has well-defined density and low thermal factors (~23 A^2 for Gln37 of the L chain vs ~32 A^2 average for the L chain and ~17 A^2 for Gln37 of the R chain vs ~26 A^2 average for the R chain), suggesting that the water-mediated interactions restrain the position of the sidechain. The position and rotameric form of Gln37 are such that the sidechain methylene group contacts the Cγ of theVal33 (3.93 Å). Similarly the methylene group (Cβ) of Ser36 also contacts the Val33 sidechain (4.19 Å). The Oγ of Ser36 forms two hydrogen bonds, to the Nε (2.97 Å) and NH2 (3.10 Å) of Arg40. The NH1 of Arg40 in turn forms a hydrogen bond to the Oε2 of Glu42 (3.24 Å).

Protein-Phosphate Contacts

Each monomer of the P22R NTD forms three salt bridges (hydrogen bonds involving cationic sidechains and phosphate oxygens) and nine additional hydrogen bonds with phosphate oxygens of DNA^{9T}, giving 12 direct polar contacts with phosphate oxygens (Figure 2.4). With one exception, all protein to phosphate contacts correspond with those identified by a phosphate ethylation interference (*26, 75*). Salt bridges from NH2 of Arg11 to the O2P at position 3 [2.9 Å, T(4)] and from NH1 of Arg14 to O2P at position 2 [3.2 Å, T(3)] help anchor helix 1 to DNA^{9T}. The recognition helix engages in four phosphate contacts. The Oγ of Ser36 contacts an O1P at position 3 [2.6 Å, T(4)], the Nε of Trp38 contacts an O1P at position 7 [2.8 Å, C(33)], the NH1 of Arg40 contacts an

O1P at position 4 [3.4 Å, T(5)] , and Ser31 makes contacts as described above. The only phosphate contacts from within helix 2 are made by Gln21, which contacts the O1P of position 2 [2.8 Å, T(3)] via the backbone nitrogen, and the O1P at position 3 via ε-NH_2 [2.9 Å, T(4)]. The final sidechain to phosphate contact occurs between δ-NH_2 of Asn49 in helix 4 and the O2P at position 7 [2.8 Å, C(33)].

The backbone NH of Ser31 makes a hydrogen bond to the O1P at position 6 [3.0 Å, T(34)], while within helix 4 the backbone NH of Asn46 makes a contact with the O2P at position 7 [2.7 Å, C(33)]. A contact of the backbone NH of Glu44 with the O1P at position 8 [3.0 Å, T(32)] is the only direct protein interaction with DNA^{9T} outside the first seven bases of the operator. This contact, as shown in Figure 7, appears to position the carbonyl of the Glu44 sidechain within proximity (< 5 Å) of the O2P of thymine at position 8 [T(32)]. High negative charge density would result from the proximity of phosphates and anionic sidechains. The negative charge may sequester cations into the minor groove, and be a factor in the induction of the B' state of the non-contacted region.

Protein-Water-DNA. Eighteen water molecules form bridges between the repressor and the DNA^{9T}. Three of these waters link protein sidechains to DNA bases, via hydrogen bonds. Ordered water molecules are also associated with the α-helices 1 and 2 of the repressor that do not bridge protein to DNA. The prevalence of these protein-associated ordered solvent molecules decreases towards the center of the complex.

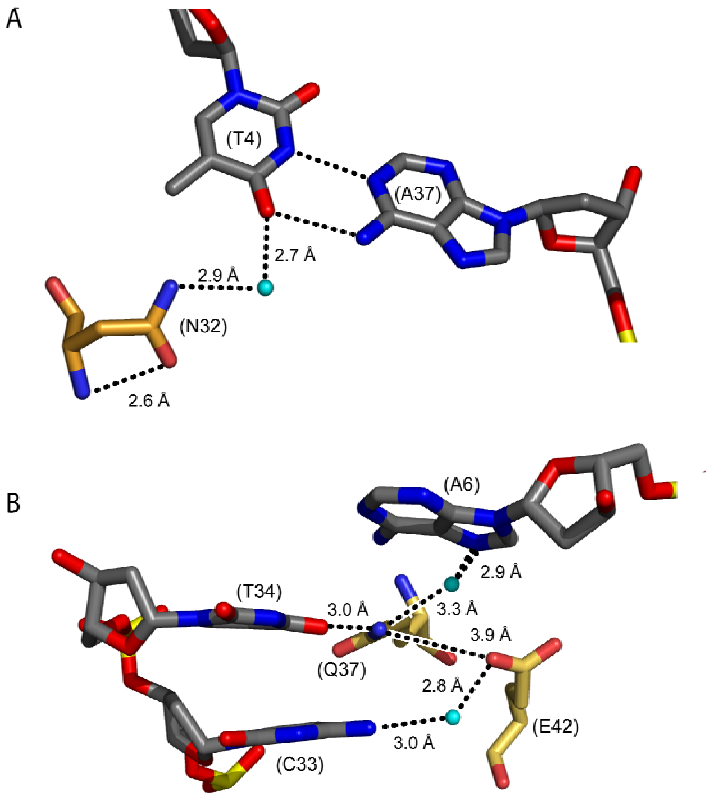

Figure 2.9. Base specific hydrogen bond interactions. (A) Water mediated interaction between Asn32 and O4 of thymine at position 3 bridged by water (cyan sphere). (B) Network of contacts between Gln37 and the N7 of adenine at position 5 bridged by water, Gln37 with the O4 of thymine at position 6, the contact between Gln37 and Glu42, and the water mediated contact between Glu42 and the N4 of cytosine at position 7.

Sidechain-Water -Base

A water molecule (water 24) links the δ-NH_2 of Asn32 and the O4 of thymine at position 3 (Figures 2.4 and 2.9A). Both the sidechain of Asn32 and water 24 are defined by clear electron density and low thermal factors ($B^{W(24)}$ = 26 A^2). The orientation and rotameric state of Asn32 appears to be stabilized by a intramolecular hydrogen bond between the δ-O and the backbone NH. Water 6 mediates the interaction between Nε of Gln37 and the N7 of adenine at position 5.

A network of interactions between Gln37, Glu42, water molecules and the cytosine at position 7 (Figure 2.9B) predicts that the base recognition at position 7 is dictated by both residues 37 and 42. The water-mediated contact between Glu42 and N4 cytosine at position 7 is the only observed base contact at this position, and may influence the identity of the base pair at that position. We speculate that the replacement of cytosine at position 7 with a purine would likely cause a change in optimal distance and geometry for the water-mediated interaction of Glu42 and position 7, resulting in the loss of affinity observed in the solution studies (*35*). However, because the water that mediates the interaction at this position could accept a proton from the N4 cytosine at position 7, as with this structure, or donate a proton to the O4 of thymine, a thymine substitution at this position is tolerated, but with a significant decrease in affinity (*35*).

The decreased affinity associated with a GC→AT substitution at position 7 may arise from a steric clash between the 5-methyl group of a thymine at position 7 and the sidechain of Glu37. The steric interference may result in a change in the position of Glu37 and cause a disruption of the hydrogen bond at position 6. Finally, Glu37 and Gln42 may be in close enough proximity (3.9 Å) to interact electrostatically with each other. The removal or change in position of Gln37 would effect the positioning of Glu42 and may result in a less stable water mediated interaction with cytosine at position 7. Currently, there are no solution studies that have tested the influence of Glu42 on base preference at position 7 and further work will need to be done to decipher this relationship.

Protein-Protein

All contacts between the two protein monomers are made by helices 4 and 5. Each monomer contains a hydrophobic patch consisting of residues Leu50, and Leu51 of helix 4 and residues Pro61 and Leu65 of helix 5. The hydrophobic patches are juxtaposed to form the dimer interface. The interface is locked on each side of the hydrophobic patch by salt bridges formed between Lys55 and Asp62 residues located on opposing monomers.

Lattice contacts

The P22R NTD – DNA^{9T} complex crystallized in the $P4_3$ space group. The DNA fragments are aligned end to end with base pairs of one duplex stacking on the base pairs of the next operator. The stacking arrangement creates a continuous nicked DNA duplex. The most disordered region of the DNA occurs at the terminal base pair at the transition from one duplex to the next. The stacking of the DNA strands is facilitated by interactions with a repressor monomer and the operator of a symmetry related complex.

In addition to the DNA-protein contacts within the complex, there are several lattice contacts, generated by symmetry. The L-chain interacts with the major groove. The R-chain interacts with the minor groove. Both of these contacts are with the R-end of the operator. There are no lattice contacts with the L-end of the operator. The asymmetry of these lattice contacts may explain the higher overall thermal factor of the L chain and half site compared to the R chain of the repressor and R half site.

The L chain of the repressor is positioned over the major groove of a symmetry-related operator, forming a direct contact of ε-NH_3 of K18 and O6 of guanine at position 7R (sym1). The R chain is positioned over the minor groove of a symmetry related

operator, allowing a sidechain-base contact of ε-NH_3 of K18 to the N3 of adenine at position 8R (sym2), creating the only break in the primary layer of hydration in the operator. In addition to the sidechain-base contacts, several salt-bridges link lysines/arginines and symmetry-related phosphates of the DNA help maintain the crystal lattice.

DISCUSSION

Mechanism of Indirect Readout

Induction of B' DNA within a Minor Groove Tunnel.

A universally-accepted definition of 'indirect readout' has yet to be established. Here we define indirect readout as the sensing of DNA sequence at a distance, in the absence of direct or solvent-mediated protein-base interaction. Haran's definition of indirect readout as DNA 'nearest-neighbor non-additivity' (*10*) is also a reasonable definition.

P22R NTD binds to DNA^{9T} in the major groove, with helices 2 and 3 forming a helix-turn-helix (HTH) motif. Central non-contacted bases of the operator modulate the affinity of the repressor for the DNA via a phenomenon known as indirect readout (*5*). The minor groove of this non-contacted region (positions 8R, 9R, 9L and 8L, sequence 5' ATAT 3') is acutely narrowed (Figures 2.5 and 2.6). The water molecules within it form a 'spine of hydration'. This observation of the B'-state in the non-contacted region of DNA^{9T} provides general predictive model of indirect readout.

DNA is known to be polymorphic, with the ability to adopt a variety of states including A, B, B' [reviewed by Hud and Plavec (*36*)] and Z-DNA [reviewed by Rich(*45*)]. The B'-state of DNA (*76, 77*) differs from the standard B-DNA by (i) a

preference for sequence elements of at least three contiguous ApA or ApT steps (5' to 3', TpA steps not favored, G-C base pairs not allowed), (ii) a propensity to bend DNA when appropriately phased (*78-80*), (iii) a narrow minor groove (*81, 82*), (iv) a spine of hydration (*37, 39*) and higher-order hydration within the minor groove(*58, 83*), (v) monovalent ions within the minor groove (*38, 53, 55, 56, 84*), (vi) negative propeller twisting(*85, 86*), (vii) unusual rigidity(*87, 88*), and (vi) a 'pre-melting' thermal transition that is distinct from duplex melting(*89-91*). The B' helical axis is thought to be linear (*92, 93*); axial bends arise at the junctions between B' and flanking B-DNA (*36*). B' hydration, especially within the minor groove, appears to be cooperative, with a many interdependent solvent interactions.

P22R NTD induces the B' state in the non-contacted (non-A-tract) region of DNA^{9T}.We propose that via this transition, the sequence of the non-contacted region is sensed at a distance, modulating the affinity for repressor. The induction of the B'-state appears to be a common feature of DNA-protein complexes. Although not previously described as B'-DNA, we observe narrow minor grooves and elements of spines of hydration in non-A-tract, non-contacted regions of DNA complexes of (i) NarI [1ZG1 (*94*)], (ii) mating-type protein α-1 [1LE8, 1AKH, 1YRN (*95-97*)], and (iii) MEF2A [1EGW(*98*)]. The induction of the B'-state may allow for optimal contacts between the protein and the operator, favorable hydration and cation distributions and electrostatics.

The Tunnel

In association with P22R NTD, the non-contacted region of the DNA^{9T} forms a "tunnel" (Figure 2.7). Solvent appears restricted in its translations within the tunnel, and could enter and exit only via the openings at the tunnel termini. Our survey of the

structural database reveals that minor groove tunnels are a unifying structural theme in complexes between DNA and lambdoid family repressors, cro dimers and other HTH proteins(*32-34, 99-102*).

Localization of Counterions in the P22 Complex

Biochemical results are consistent with localization of cations within the P22R NTD – DNA^{9T} tunnel, and support a role for cations in modulating the DNA-protein stability (23, 31, 32). The 434 repressor, but not P22R NTD, positions arginine sidechains at the lip of the minor groove, directly above the center of the non-contacted region. These cationic sidechains contribute to stability of 434 repressor DNA complexes (*5*).

The cationic arginine residues of 434 may be replaced by cationic counterions in the P22R NTD - DNA^{9T} complex. The P22R NTD positions two negatively charged glutamate residues (Glu44 and Glu48, Figure 2.7) in the ceiling of the tunnel, in close proximity to DNA phosphate groups (4.9Å and 7.9Å, respectively) as well as to each other (4.4 Å). The localization of monovalent cations in the minor groove of the P22R NTD-DNA^{9T} complex is consistent with the induction of the B'-state of DNA^{9T} in the tunnel region. Monovalent cations are known to localize in the minor groove of B' DNA(*38, 53, 55, 56, 84*). The tunnel has a high electronegative potential which seems likely to result in cation localization.

The Mechanism of Direct Readout

The Valine Cleft

The structure of the P22R NTD – DNA^{9T} complex suggests that sequence-specificity arises substantially from interaction of Val33 with a complementary binding

cleft on the major groove surface of DNA^{9T} (Figure 2.8). The cleft is formed by four methyl groups on sequential base pairs (5' TTAA 3'). This cleft is intrinsic to the DNA sequence and is not induced by P22R NTD binding. The importance of the Valine Cleft is consistent with the influence of Valine 33 on the base pair preferences at positions 3 - 6 (*29*).

To our knowledge, descriptions of this or related clefts within the major groove are not found in the literature. However hydrophobic DNA-protein contacts have been noted in structures such as DtxR(*103*), λ cro (*101*) and repressor (*104*), IdeR (*105, 106*), NarL(*107*), PhoB(*108*) and E. coli replication terminator protein(*109*). The importance of hydrophobic interactions in DNA-protein affinity and recognition have been discussed (*103, 105, 110*).

Hydrogen Bond Interactions

Hydrogen bonding of Gln37 contributes to the requirement of an AT base pair at position 6. The presence of a Gln at position 37 is necessary for binding of P22R NTD to the operator (*35*). This requirement is consistent with the interaction between Gln37 and thymine at position 6 (the only direct sidechain-base hydrogen bond in the complex, Figures 2.4 and 2.9). In a secondary contribution, hydrogen bonding may restrain the sidechain of Gln37 in such a way that the aliphatic groups of the sidechain contribute hydrophobic interactions with Val33.

The requirement for a GC base pair at position 7 is not dictated by simple directional interactions. P22 repressor is intolerant of substitutions at position 7 other than G-C to A-T, which reduces P22R affinity by a factor of 55 (*35*). This G-C requirement at position 7 is highly influenced by Gln37, although no direct contacts link

Gln37 with position 7. It appears that subtle solvent organization, shape-complementarities and long-range electrostatic interactions dictate this specificity.

The water mediated interactions of Asn32 and the O4 of thymine 3 has been indicated by biochemical data to contribute to the necessity of T•A base pair at this position. The specificity of this interaction is somewhat surprising since the contact is water mediated and because a hydrogen bond donor and a hydrogen bond acceptor are available from Asn32. The water molecules that mediate the contacts, one on each end of the operator are well ordered, judging from their well-defined density and thermal factors. A comparison of the NMR model of the free repressor and the crystal structure of the bound repressor reveals very little conformational change in the unbound and bound states (*31*). However, the hydrogen bond between δ-O of Asn32 and the backbone NH of the residue is not observed in the NMR of the free repressor. The recruitment of the water molecule to bridge the interaction may optimize the geometry for the hydrogen bond between δ-O of Asn32 and the backbone NH of the residue, giving rise to the specificity of the interaction and the intolerance of mutation at position 3 observed in biochemical studies (12).

CONCLUSION

We have found that sequence-dependent DNA shape plays a role in direct readout of DNA sequence by P22 repressor. The binding of the P22R NTD to DNA^{9T} appears to be contingent on a complementary cleft for Val33 formed by arrangement of 5-methyl groups of thymines at positions 3,4,5 and 6.

We have also proposed a model by which sequence-dependent DNA structural differences modulate protein-DNA interactions of not only P22 repressor, but also all

lambdoid family repressor and cro proteins, as well as other HTH family proteins. In our structure of the P22R NTD-9T complex we find that solvent molecules are recruited into a tunnel composed the dimer interface and the minor groove of the non-contacted region of the operator. We believe that the sequence of bases in this area allows for a difference in affinity for monovalent cations, similar to that of cationic sidechains of 434 repressor and cro and λ repressor and cro. This difference in affinity for cations would result in different binding affinities that are dependent on the base content in the non-contacted region. Given the similarities in the P22R NTD/9T complex with other HTH complexes the mode of differential binding may extend to other HTH complexes outside the lambdoid family, as well as to complexes not yet investigated.

As with most DNA binding proteins, the affinity of the P22R NTD to for its operator is not simply accomplished by the recognition of DNA functional groups by protein sidechains. Although this direct recognition is the most recognizable aspect of protein-DNA complex formation, it is only one of the components that are integrated with DNA conformation which is dictated by the sequence, and DNA deformability which is dictated by how a particular sequence responds to interactions with both protein and solvent molecules. The further categorizing and characterizing of DNA-protein interactions individually as well as how they work in concert with each other is key to understanding and predicting interactions of DNA with proteins and other molecules.

CHAPTER 3

P22 C2 REPRESSOR-DNA^{9C} OPERATOR COMPLEX

The discrimination of the P22 c2 repressor for different operators is dependent on both direct and indirect methods of DNA recognition (*23, 28, 29, 35*). The previous x-ray structure of the P22R NTD in complex with a synthetic operator, DNA^{9T}, suggests that direct recognition of the DNA sequence occurs by the insertion of recognition helix into the major groove of the outer seven bases on each end of the operator, with no direct base contacts occurring between the central four bases and the repressor.

Binding assays from Koudelka have shown that the P22 c2 repressor has a 10-fold lower affinity for a synthetic operator that contains a CpG step at the central non-contacted region of the sequence, DNA^{9C}, compared to that with a T-A step at the central base step, the DNA^{9T} operator (*28*). Furthermore, the operators differ in the salt-dependence of their affinities (*29*). We have proposed that the optimal binding of the P22R NTD occurs when the non-contacted DNA region is in the B' state, which is characterized by a spine of hydration and an unusually narrow minor groove. Thus, the indirect readout mechanism of the P22 c2 repressor occurs due to an induced transition of B-DNA to B'-DNA. The propensity associated with a B→B' state transition is known to be sequence dependent, and providing a predictive model of indirect readout based on the sequence of the non-contacted DNA.

The mechanism of indirect readout allows for several possibilities to account for the change affinity due to a mutation in the non-contacted region. The interaction of the P22R NTD with the operator could stabilize a B' state. However, the energy cost associated with the induction of the B' state of the operator (i.e. spine of hydration,

narrow minor groove, propeller twisting, and the localization of cation within the groove) is higher for DNA^{9C} than that for DNA^{9T}. The higher energy cost associated with the protein induced DNA conformation would result in a less stable conformation for the DNA^{9C} complex.

Alternatively, the energy cost is too high to obtain the narrow minor groove of the B' state in the presence of a CpG step at the central bases. The high energy cost may result in an increase minor groove width, a change in the helical parameters, and/or a disruption in the contacts between the repressor and the operator for the DNA^{9C} complex.

The x-ray crystal structure of the P22R NTD with the synthetic operator DNA^{9C} presented here demonstrates that indirect readout is indeed associated with the transition to the B→B'state. A comparison between the DNA^{9C} and the DNA^{9T} complexes reveals that all contacts are conserved between the two structures. Also, the DNA helix is unaffected by the TpA→CpG mutation. The solvation of the two complexes is also conserved, including the spine of hydration in the non-contacted region. However, the crystal structure of the P22R NTD-DNA^{9C} complex provides evidence that the difference in the stability of the DNA^{9C} and the DNA^{9T} structure arises from the difference in the stability of the spine of hydration in the minor groove of the non-contacted region.

Materials and Methods

Construction of plasmid expressing P22R NTD

A DNA fragment encoding the NTD fragment of P22 repressor was created using PCR to insert a stop codon at position 69 of the intact P22 repressor gene. We amplified a 450 bp fragment from pTP125 (*64*) using primers obtained from the CAMBI Nucleic Acid Facility (University at Buffalo). The sequences of the primers were:

5'GTTTTTTGCGTCGACATCATA 3' and 5' CGTTTGTCTGAC GTCAATCTCCTTTC 3'. This fragment, containing the tac promoter (*65*) fused to the P22 repressor gene fragment, was cleaved with Aat II and Sal I, gel purified, ligated in pUC18 cut with the same enzymes and transformed into E. coli strain XA90 (*66*) . The resulting plasmid expresses a protein comprised of the first 68 amino acids of P22 repressor.

Purification of P22R NTD

E. coli XA90 cells (*66*) bearing P22R NTD expression plasmid were grown with aeration at 37^{o}C in 3 liters of Luria broth supplemented with 0.1 mg/mL ampicillin to an A_{600} of approximately 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mM, and growth at 37^{o}C was continued for 4 hours. The cells were harvested by centrifugation and resuspended in 20 mL lysis buffer (200 mM NaCl, 50 mM EDTA; 100 mM Tris pH 7.4) with 0.5 mL of 20mg/mL phenyl methyl sulfonyl fluoride (PMSF). The cells were lysed in a French pressure cell (18,000 lb/in^2) at 4^{o}C. The lysate was diluted with 100 mL of cold lysis buffer. Cellular debris was precipitated by centrifugation at 4000 x g. Nucleic acids were precipitated by dropwise addition of polyethyleneimine to a final concentration of 0.6% at 4^{o}C and mixed for 15 minutes, followed by centrifugation at 10,000 x g. Soluble proteins were precipitated with 40% (w/v) ammonium sulfate and dialyzed at 4^{o}C against three 1 L changes of standard buffer (SPB-50 mM $NaPO_4$ pH 6.8 10 mM EDTA) supplemented with 50 mM NaCl. The dialyzate was loaded onto three Econo-Pac carboxymethyl (CM) cartridges (Bio-Rad, Hercules, CA) and eluted with a linear salt gradient from 0.1 to 1 M NaCl in SPB over 70 minutes. Fractions containing repressor were pooled and precipitated with 40% ammonium sulfate. Following

precipitation, the protein-containing pellet was dissolved in SPB+100 mM NaCl and loaded onto a 1 cm X 50 cm Sephacryl S100 (Amersham Biosciences) size exclusion column. The column was developed with the same buffer at a flow rate of 0.3 mL/min. Fractions containing P22R NTD were pooled and the protein was precipitated with 40% ammonium sulfate. The dissolved pellet was dialyzed against SPB supplemented with 150 mM NaCl and 20% glycerol. The protein was flash frozen in liquid nitrogen and stored at -70°C.

DNA purification

DNA oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA) as an ammonium salt from reverse HPLC purificaton. To form the double-stranded protein binding site, equimolar amounts of the complementary strands were mixed in water, heated to 85°C for 60 s and slow-cooled over four hours to anneal.

Co-crystallization of P22R NTD with DNA^{9C}

Crystals were grown at 4°C by the hanging drop vapor diffusion method. The initial crystallization solution contained 0.42 mM P22R NTD, 0.42 mM duplex d($^{5'}$TATTTAAGACGTCTTAAATG$^{3'}$) -d($^{5'}$CATTTAAGACGTCTTAAATA$^{3'}$), 45 mM Tris·HCl (pH 7.8), 20 mM NaCl, 2.0 mM glycerol, 14% PEG 400, 4.5 mM LiCl, 2.3mM $MgCl_2$ and 0.9% MPD in a volume of 5.3 μl. The crystallization solution was equilibrated against a reservoir of 100 mM Tris·HCl (pH 7.8), 30% PEG 400, 10 mM LiCl, 5 mM $MgCl_2$ and 2% MPD. Crystals grew to a size of approximately 0.5x0.3x0.3 mm^3 within one day. A single crystal was collected from the crystallization solution in a nylon loop and flash frozen in liquid nitrogen.

Data Collection

Diffraction data were collected on beamline X26-C at The National Synchrotron Light Source, Brookhaven National Lab with an ADSC Quantum 4 CCD detector using radiation of 0.99997 Å wavelength. The crystal was maintained at 110 °K. A total of 232,653 reflections were indexed and integrated with DENZO (*67*) and reduced to 32,768 unique reflections in space group $P2_12_12_1$ in the resolution range of 35-1.60 Å. Data were scaled with the program SCALEPACK (*68*).

Refinement

Initial phases were determined by molecular replacement with the program Molrep of the CCP4 suite (*111*) using data from 15 Å to 3Å. The P22NTD-DNA^{9T} complex with waters removed was used a search model. Following a rigid body refinement in Refmac5 (*112*), difference density was observed in a the $|F_o-F_c|$ maps at positions 9L and 9R of the operator corresponding to T-A→ C-G mutations at this position. The data were extended to 1.67 Å, and a C-G base pair was modeled into positions 9L and 9R based on $|2F_o-F_c|$ and $|F_o-F_c|$ maps using the program COOT (*113*), followed by refinement in Refmac. Further model building was performed by fitting proper protein sidechain rotomers in COOT. Water molecules were added iteratively to the model initially using ArpWarp and by inspection of $|2F_o-F_c|$ and $|F_o-F_c|$ maps in later stages, followed by cycles of refinement in Refmac, resulting in a R-factor of 22.25% and a R_{free} of 25.59%. Two final rounds of TLS refinement was performed in Refmac using parameters shown in the appendix. The final refined model of the P22R NTD (residues 1-68) in a complex with DNA^{9C} is refined to 1.67 Å resolution, with a R-factor 19.78 % and a R_{free} of 22.58 %. Data collection and refinement statistics are given in Table 3.1.

Table 3.1. Data Collection and Refinement Statistics

X-ray source	X26-C National Synchrotron Light Source
Wavelength (Å)	0.99997
Detector	ADSC Quantum 4 CCD
Resolution Range (Å)	19.90-1.67
Mosaic spread (°)	~0.7
Space Group	$P2_12_12_1$
Unit-cell parameters (Å)	
A	40.602
B	54.319
C	114.112
Total Reflections	347,031
Unique Reflections	27753
Multiplicity	5.8 (4.7)*
Completeness (%)	97.3 (91.7)
Highest Resolution Shell (Å)	1.67-1.71
$I/\sigma(I)$	47.1 (2.12)
Rsym (%)	6.9 (65.9)
Reflections (working/test set)	27753
Number of non-H protein atoms	2 x 514
Number of non-H DNA atoms	814
Number of water atoms	330
R-crystal	19.78%
R-free	22.45%
Average B-factor	24.78
RMSD from ideal	
Bond length (Å)	0.01
Bond angle (°)	1.59

* The number in parenthesis indicates the value for the highest resolution shell.

Electron Density and Model Quality

The $2F_o$-F_c Fourier maps clearly define the protein backbone and sidechains with exceptions occurring only with the N-terminal residues and a small number of solvent exposed sidechains. Multiple conformations for several sidechains were built into the model when clearly defined by both sum and difference density and using R-factor and R_{free} as guides. The N-terminal residue of the L chain is not clearly defined by density and the sidechain of residue 2 is also highly disordered. Therefore residue one has been omitted from the model and the sidechain of residue two has been built into the model at 1% occupancy. The first three N-terminal residues of chain R are undefined in sum maps and have been omitted from the model.

The electron density of the DNA is sharp and well-defined throughout with exceptions at the terminal base pairs where base stacking is disrupted between DNA stands at the junction of asymmetric units (Figure 3.1). Asymmetry occurs in the DNA oligomer at the terminal base pairs in that one of the terminal base pairs is a G·C base pair and the other is an A·T. The R-factor and R_{free} as well as the difference density indicate a best fit with the G·C base pair in the R half site terminus and the A·T in the L half site terminus. The G of the R terminal base pairs likely contains multiple conformations and occupancy. A conformation of the G at the terminal G·C base pair is likely present that optimizes Watson-Crick base pairing. However, efforts to model multiple conformations were not justified by fitting statistics. Therefore, only a single, predominant position of the G in a over-twisted conformation was modeled for this terminal residue.

A Ramanchandran plot (*71*) calculated with Procheck (*114*) shows that 93% of protein residues fall in the most favored regions of backbone conformation, while 7% of

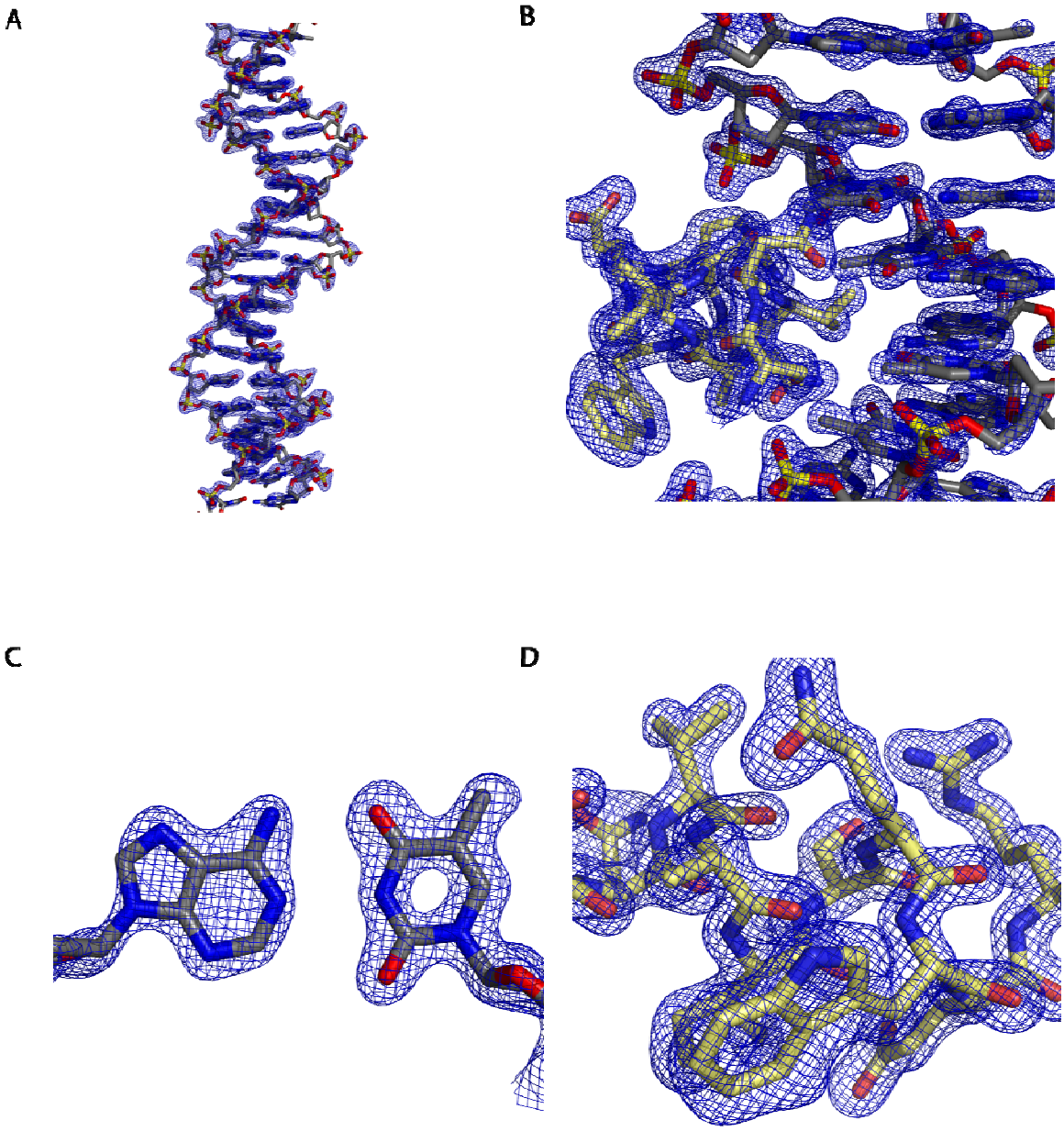

Figure 3.1. The electron density defining the DNA, protein and solvent is of high quality. Shown here is (A) electron density surrounding the DNA duplex contoured at 1.5 σ, (B) electron density surrounding minor groove solvent contoured at 1.3 σ, (C) electron density surrounding base pair T(9)-A(32) contoured at 1.3 σ, (F) electron density surrounding amino acids of the recognition helix (residues 31-40) contoured at 1.5 σ, (E) electron density surrounding the DNA and the R-recognition helix contoured at 1.5 σ. The maps are annealed/refined 2Fo-Fc Fourier synthesis maps.

residues fall in allowed regions. No residues fall in the generously allowed or forbidden region of the Ramanchandran plot.

Results

The Global Complex

The overall architecture of the P22R NTD-DNA^{9C} complex is unaffected by a CpG substitution as compared to the P22R NTD-DNA^{9T} complex. The P22R NTD binds as a dimer to the operator, with each monomer containing five α-helices. As with the DNA^{9T} structure, the P22R NTD-DNA^{9C} complex is highly symmetric about a pseudo-two fold rotation centered at position 9L and 9R of the operator. The symmetry again extends down to the solvent molecules which are clearly observable in the electron density maps.

Helix two and three form a helix-turn-helix (HTH) motif. As with the DNA^{9T} complex, helix three is the most ordered region of the complex judging from the thermal factors and the electron density maps. Helix three of each monomer, the recognition helix, is inserted into the major groove of the operator one turn away from each other. The alignment of the recognition helix position both monomers on the same face of the operator and all the formation of the dimer interface directly over the minor groove (Figure 3.2).

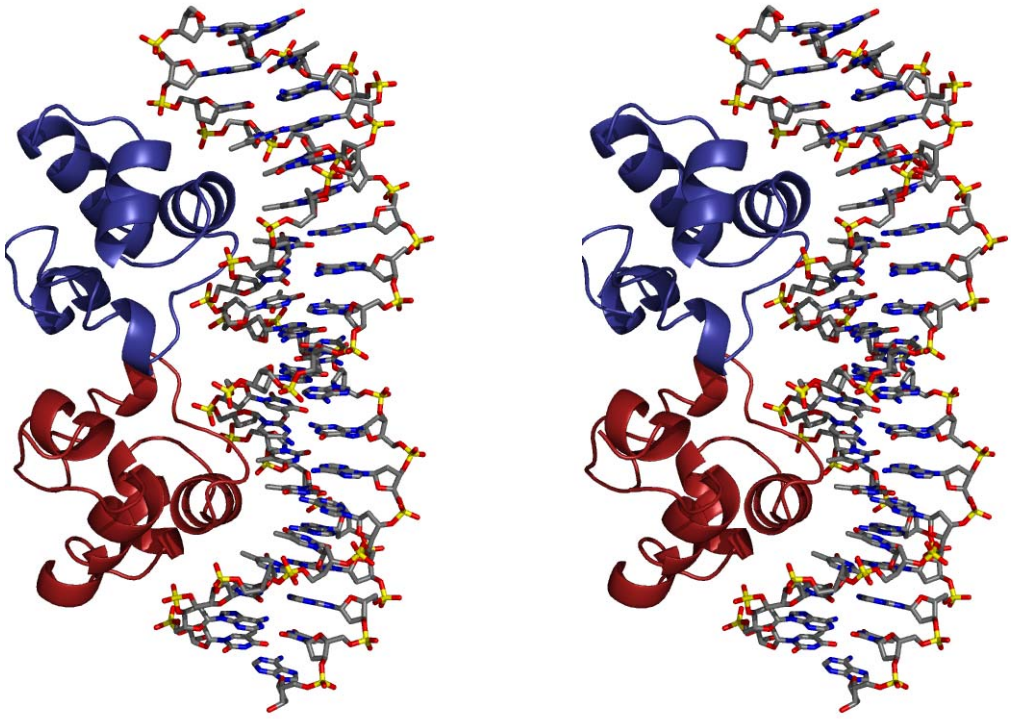

Figure 3.2. (A) Stereo view of the DNA^{9C} protein complex. The R-subunit is represented by red ribbon; The L-subunit is represented by blue ribbon. The DNA represented as stick with C, gray; N, blue; O, red; P, yellow.

Global DNA Conformation

An inspection of the helical parameters (CURVES v. 5.3) shows that the overall DNA conformation DNA^{9C} is conserved compared to that of DNA^{9T}. The RMSD of the DNA^{9C} and DNA^{9T} backbone atoms is 0.41 Å when protein backbones are superimposed. The deviation from linearity of the local axis between base pair 5 and 6 creates a kink (9.3°) in the DNA^{9C} is consistent with the kink observed at this position in the DNA^{9T} strucuture (8.8°) that facilitates DNA contacts with the recognition helix. The positive roll between base pairs 5 and 6 (9°) and between base pairs 6 and 7 (10°) are again present in the DNA^{9C}. As with the DNA^{9T} structure, the kink in the DNA^{9C} complex

occurs between the naturally occurring A-tract of the contacted region and the induced B'-DNA (see below) of the non-contacted region of the operator, and is such that the DNA helical axis wraps toward recognition helix.

The Non-contacted Region

The CpG substitution

The CpG base pair substitution in the DNA^{9C} for the TpA base pair in the DNA^{9T} structure at position 9L and 9R results in small changes of the helicoid parameters. The most significant changes are in the base pair parameters of buckle (12.7° DNA^{9C}, 6.4° DNA^{9T} at position 9L and -13.0° DNA^{9C}, -9.1° DNA^{9T} at position 9R, indicated by 12.7°, 6.4°, 9L; -13.0°, -9.1°, 9R), propeller twisting (-9.4°, -16.0°, 9L; -10.6°, -5.2°, 9R) and opening (0.49° DNA^{9C}, 2.18° DNA^{9T} at position 9R). Deviations also occur at the 9L-9R step in the twist angle (42.58° DNA^{9C}, 44.22° DNA^{9T}) and roll (0.96° DNA^{9C}, -9.94° DNA^{9T}). Despite the changes in the helical parameters between the DNA^{9C} and DNA^{9T}, the N3 of the guanine at position 9R and N3 of the guanine at position 9L of DNA^{9C} remain in the same relative position. The positioning of the N3 atoms of the purines at the 9L-9R step appear to be critical in the coordination of a solvent molecule (see below and Discussion).

The Induced B' State

The non-contacted region of the DNA^{9C} is in the B' state as defined by a spine of hydration and an especially narrow minor groove. To our knowledge the B' state has not been observed in the crystal structures of DNA sequences containing a CpG step. The B' state of the non-contacted region is induced by the binding of the repressor. The

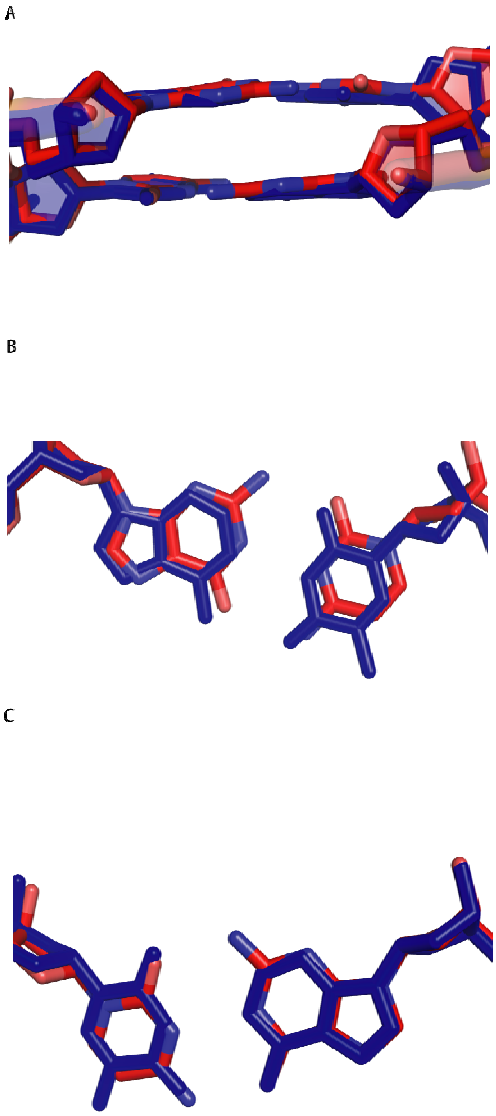

Figure 3.3. A) Superimposition of positions 9L and 9R of DNA^{9C} (red) and DNA^{9T} (blue). View is looking into the minor groove. B) Superimposed base pair of position 9L. C) Superimposed positions 9R.

conformation of the non-contacted region is similar to both the A-tract DNA sequences present in the DNA^{9C} structure here, and the non-contacted region and A-tracts of the DNA^{9T} operator. The B' state of the non-contacted region appears to be highly conserved in P22R NTD binding with RMSD of backbone atoms between the non-contacted region of the DNA^{9C} and DNA^{9T} is 0.29 Å. The B' of the non-contacted region is likely required for the proper positioning of the phosphate backbone to form hydrogen bonds with the repressor.

Minor groove hydration

The spine of hydration in DNA^{9C} is very similar to that of the DNA^{9T}, in that it is continuous from position 8L-8R. Also, the spine of hydration in the non-contacted region contains a primary layer of hydration that forms hydrogen bonds with the functional groups on the floor of the minor groove. There is also a secondary layer of hydration perched on the primary layer of hydration, and a third layer of hydration is observed which sits atop the secondary layer and contacts residues of dimer interface of the repressor in the DNA^{9C} complex.

The primary and secondary layers of the spine of hydration in the DNA^{9C} are not identical to that of DNA^{9T}. The density surrounding the water molecules in the primary and secondary of the spine of the non-contacted region of the DNA^{9C} is not well defined compared to that of the DNA^{9T} complex. The broadening of the peaks that correspond to these water molecules may indicate that these water molecules are less ordered than those in the DNA^{9T} structure or of those found in A-tract DNA^{9C}. Several lines of evidence indicate that peak broadening are a characteristic of the CpG step in the DNA^{9C} complex and does not result from differences in crystallization conditions, etc.

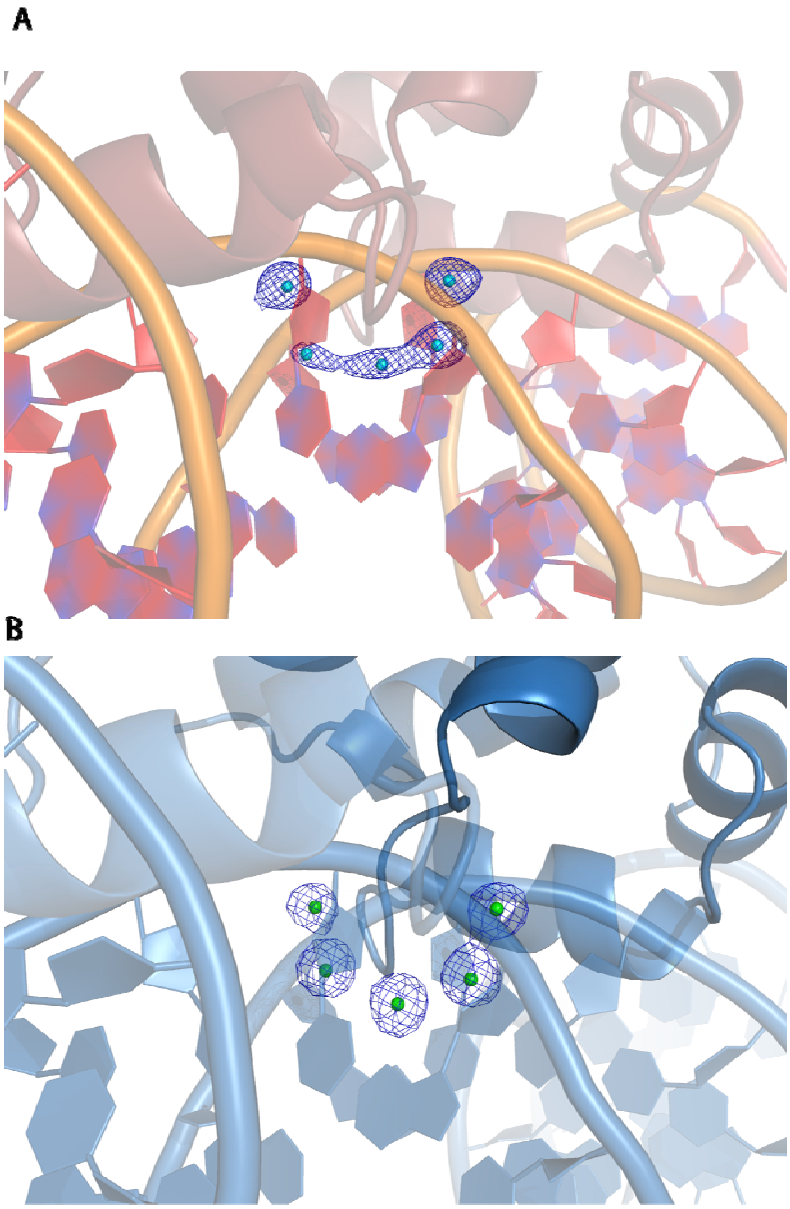

Figure 3.4. The spine of hydration of the non-contacted region in the A) P22R NTD-DNA^{9C} and the P22R NTD-DNA^{9T} complexes. The DNA is represented in cartoon, and the repressor is represented in ribbons. The density surrounding the water is the result of a $2F_o$-F_c Fourier synthesis map contoured at 1.1σ.

First, the third layer of hydration in the non-contacted region is well defined in the non-contacted region of the DNA^{9C} complex and is similar to the DNA^{9T} structure. Second, the primary and secondary layers of the spine of hydration in the naturally occurring A-tract of the DNA^{9C} complex are well defined. Finally, the P22R NTD-DNA^{9C} complex has been solved in a $P4_3$ spacegroup (data not shown) at 2.0 Å resolution, and the primary and secondary layer of the spine of hydration in the non-contacted region again shows the peak broadening observed in this structure.

The primary water molecule at the central CpG step in the DNA^{9C} complex forms N3-W103-N3 bridge between the guanines at position 9L and 9R. The water is hexa-coordinate with the two hydrogen bond accepting nitrogens of the cross-strand guanines, two water molecules of the secondary layer of the spine, and O4's of the deoxyribose sugar in the backbone. This coordination is analogous to that of the N3-W-N3 coordination at position 9L and 9R of the DNA^{9T} operator, and the O2-W-O2 coordination in A-tracts containing a ApT step. The RMSD between the primary waters of the DNA^{9C} (water 130) and DNA^{9T} of 0.36 Å is the highest difference in position of any waters within the induced B' state non-contacted region.

Narrow Minor Groove in the Non-Contacted Region

The narrow minor groove, a characteristic of the B' state, reaches a minimum at the center of the non-contacted region at position positions L9 and R9. The minor groove begins to widen going away from the center (R8 through R5 and L8 through L5). The minor groove then narrows again in the A-tracts at each end of the operator (R4 through R1 and L4 through L1).

The Tunnel

The tunnel is structurally unchanged. The DNA^{9C} tunnel, comprised of the minor groove in the non-contacted region and the dimer interface of the repressor, is electrostatically different than that of the DNA^{9T} complex. The floor of the tunnel is comprised of the base edges of residues 11-13, and 31-32. The A→G substitution at base 11 and base 31 introduce the positive potential of the N2 exo-cyclic amine of guanine to the floor of the tunnel. The walls of the tunnel in the DNA^{9C} structure, made up of the sugar-phosphate backbone, are unperturbed relative to their position in the DNA^{9T} structure. The ceiling of the tunnel formed by the dimer interface of the repressor, is made up from the loop region connecting helices 3 and 4 and the N-terminal region of helix 4.

The tunnel is filled with ordered water molecules, and as in the DNA^{9T} structure, the presence of the tunnel in the DNA^{9C} structure allows movement of water molecules or cations into and out of the tunnel by moving along the floor of the minor groove. As discussed above, an ordered water molecule interacts with the floor of the tunnel to form the primary layer of the spine of hydration. In the ceiling of the tunnel, the δNH_2 of Asn46 on each of the repressor monomers forms hydrogen bonds with the tertiary waters of the spine in the tunnel. These interactions are conserved compared to the DNA^{9T}, and show the lowest deviation in position of any of the waters in the induced spine of hydration in the tunnel.

DNA-Protein contacts

The DNA-protein contacts of the P22R NTD-DNA^{9C} complex are conserved compared to those of the P22R NTD-DNA^{9T} complex (Appendix). The RMSD of all

protein backbone atoms is 0.59 when the DNA phosphodiester backbones of the DNA^{9C} and DNA^{9T} operators are superimposed. The conservation of DNA conformation, protein positioning and contact distances is evidence that there is virtually no change to the complex due to the CpG mutation at positions 9L and 9R outside of the tunnel.

As in the DNA^{9T} complex, the direct recognition of the DNA^{9C} operator occurs due to interactions of amino acid residues 31-33, 37 and 42, which are within or directly flanking the recognition helix (Helix 3). The sidechains of these residues are restrained relative to the DNA (i.e. when DNA backbones of DNA^{9C} and DNA^{9T} are aligned) with RMSD's of 0.10 Å (Ser31), 0.09 Å (Asn32), 0.09 Å (Val33), 0.10 Å (Gln37), and 0.37 Å (Glu42) for all atoms in the sidechain. Residues 31, 33, and 37 make direct sidechain-base contacts. In doing so, these protein residues help establish the discrimination between the base pairs at position 3-6 of the operator by the repressor (Figure). Asn32 forms a water-mediated interaction with the O4 the thymine at position 3 in the DNA^{9C} complex and in the DNA^{9T} complex. Solution experiments have indicated a highly specific interaction between Asn32 and thymine at position 3 of the operator.

It is likely that a water mediated contact with Asn32 does impose the high degree of discrimination at position 3 of the operator indicated by the solution experiments. The specificity of this interaction is surprising due to the diverse possible orientations of the hydrogen bonding functionalities of water and Asn. The orientation of the hydrogen bonding network is likely stabilized by a intramolecular hydrogen bond of Asn32 between the δ-O and the backbone NH. The high degree of conservation in the contact distances, the positions of the atoms in the Asn32 sidechain, and the conserved position at each end of the operator of the water that mediates the interaction indicate that this

interaction is not the result of lattice contacts, experimental condition, or errors in data processing.

The Valine Cleft

As with the DNA^{9T} complex, valine 33 is inserted into a prominent cleft in the major groove of the DNA^{9C}. The Valine Cleft is composed of two opposing sets of major groove methyl groups contributed by A-T base pairs of the sequence 5' TTAA 3'. The 5-methyl groups of positions 3 and 4 form one hydrophobic wall with contributions from the methyl group of the T-A at position 2. The 5-methyl groups of positions 5 and 6 form a second hydrophobic wall. The cleft appears to be a natural consequence of the TTAA sequence element and is not induced by protein binding. An empty cleft with similar dimensions is observable in an unbound DNA fragment with the same 5' TTAA 3' sequence element (NDB entry BD0051 (*72*)) has been described previously (see chapter 2) and is shown in Figure.

The shape complementary to that of the valine sidechain is conserved and the position of the Val33 is restrained compared to the DNA^{9T} complex (Figure) with all atoms involved in the cleft superimposing with an RMSD of 0.086 Å. The Val33 interaction with the valine cleft is a unique binding element to the P22 c2 repressor and has not been observed in other structures. Interactions of Val33 within this pocket appear to be important for positioning the protein relative to the DNA, and may confer thermodynamic stability and specificity.

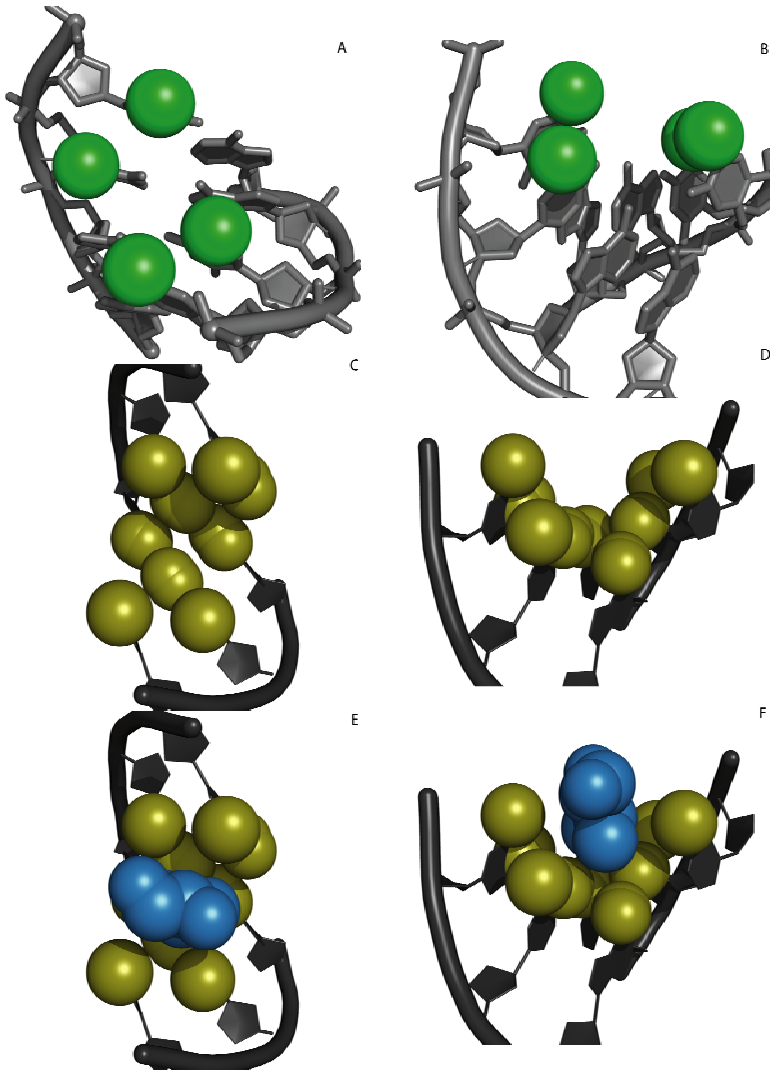

Figure 3.5. The Valine Cleft. The cleft of a TTAA found in unbound DNA. A)The view is looking into the major groove, B) and looking along the major groove. The grey tube represent the DNA backbone, and the green spheres represent the methyl groups of the thymines that make up the hydrophobic region of the cleft (PDB ID: 167D) Panels C-F) The surface of the DNA atoms composing the R-half site Valine Cleft is represented in A green tube highlights the backbone. The functional groups of the major groove that form the Valine Cleft are gold. Val 33 is shown in space filling representation is colored blue. (C) DNA only, viewing into the major groove, showing the 'empty' Valine Cleft. (D) Same representation as panel C except viewing along the major groove. (E) The DNA plus Val 33, showing the shape complementarity of the Val 33 and the Valine Cleft. (F) Same representation as panel E except viewing along the major groove.

Protein-Phosphate Contacts

The P22R NTD forms six salt bridges (three per monomer) and 18 additional hydrogen bonds (nine per monomer) with phosphate oxygens of the DNA^{9C} operator. These contacts are the same as previously described for the DNA^{9T} complex and are shown in Appendix, giving 12 direct polar contacts with phosphate oxygens. We concentrate here on the protein-phosphate contacts that occur within the tunnel region to confirm the consistency of the contacts in the region of the TpA→CpG mutation (Figure).

As with the DNA^{9T} complex, the δ-NH_2 of Asn49 of the L monomer in the DNA^{9C} complex contacts the O2P at position 7 [2.8 Å, C(33)]. The backbone NH of Asn46 contacts the O2P at position 7 [2.7 Å, C(33)]. The Nε of Trp38 contacts the O1P at position 7 [2.8 Å, C(33)]. The backbone NH of Glu44 contacts the O1P at position 8 [3.0 Å, T(32)]. The position of the carbonyl of the Glu44 sidechain remains consistant with those of the DNA^{9T} complex, and are within proximity (< 5 Å) of the O2P of thymine at position 8 [T(32)]. These phosphate contacts are mirrored by contacts of the R monomer of the repressor dimer in both location and contact distances. The contacts between the repressor with the phosphates in the tunnel are complimentary to the B' state, indicating that any deviation from the B' state would result in a less stable complex.

Discussion

Effects of GC Substitution on Direct readout

In the structure of P22R NTD in a complex with operator DNA^{9C} we observe that the overall architecture of the complex is conserved compared to the complex with DNA^{9T}. The superimposition of the structures based on all protein and DNA backbone atoms gives an RMSD of 0.28 Å. The position and interactions of the recognition helix, helix 3, is conserved, giving an RMSD of atomic positions of 0.22 Å for all atoms from residue 31 to 40.

B' state

We have proposed a model in which indirect readout by the P22 c2 repressor is caused by the induction of the B'-DNA state in the non-contacted region of the DNA (*115*). The differential energy cost of this transition is associated with the differential propensities of various DNA sequences to adopt to B' state. The stability of the B'-state varies with sequence (ApA, ApT > TpA > CpG, CpC, GpC, GpG) and length. The effect of the sequence of the non-contacted region on the stability of the complex is consistent with the importance of the B' state. Sequences that are most resistant to the B'-state (*36*) have lowest affinity for P22R (23, 30). For example CpG base pairs within the non-contacted region show the greatest destabilizing effect.

Intrinsic properties of the C•G base pair

The structure presented here of the P22R NTD in complex illustrates, for the first time that a sequence with a CpG step can adopt the B' state. We propose that while the non-contacted region of the DNA^{9C} is capable of establishing the B' state, it is at a higher

energy state than the DNA^{9T} induced B' state. The CpG step at position 9L and 9R show reduced propeller twisting and greater buckling compared to the TpA base pair step of the DNA^{9T} structure, as well as reduced opening. The difference in intrinsic properties of the base pairs at the central step can not be ignored. The O2-NH5 hydrogen bond of a C•G base pair impairs the propeller twisting and base pair opening that are likely necessary to conform to the B' state. The reduction in propeller twisting and base pair opening are likely compensated for by the increase in base pair buckle observed at the CpG step in DNA^{9C} compared to that of the TpA step in the DNA^{9T}. However, the loss of base stacking and the strain of the minor groove hydrogen bond likely results in a less stable conformation for the CpG step.

Minor Groove width

The minor groove width of the DNA^{9C} operator is narrow in the non-contacted region compared to B form DNA. The reduced width of the minor groove is similar to that of the B' state observed in naturally occurring A-tracts and the induced B' state of the non-contacted region of the DNA^{9T} operator in complex with the P22R NTD. The narrow minor groove observed in the P22R NTD-DNA^{9C} complex contains a CpG step. All structures, both x-ray and NMR, of free DNA show a wider minor groove that is similar to B-form DNA or A-form DNA when a CpG step is present(*36, 85, 116*). The narrow minor groove is induced by the binding of the P22R NTD. The positions of the phosphates of narrow minor groove in the induced B' state are complimentary to the hydrogen bond donating functional groups on the DNA face of the repressor.

Solvation

We observe a spine of hydration extending through the minor groove of the non-contacted region of the DNA^{9C} operator. The spine of hydration present in A-tracts is essential for the adoption of the B' state. The solvation in the non-contacted region of the minor groove of the DNA^{9C} structure remarkably similar to that of naturally occurring A-tracts and the non-contacted region of the DNA^{9T} operator. The solvent molecule at the central step is coordinated by cross-strand interaction with the N3 of a purine (G11 and G31 of DNA^{9C} and A11 and A31 of DNA^{9T}). The coordination of the water molecule at the central step of both operators is analogous to the O2-W-O2 and O2-W-N3 coordination of A-tracts. The non-contacted minor groove of DNA^{9C} and of DNA^{9T} both have higher ordered water that extend to the ceiling of the tunnel.

A close inspection of the electron density of the water molecules at the central position of each operator, however, shows that the density of the primary water in the DNA^{9C} operator is much less defined than that of the water in the DNA^{9T}. The density of the water molecule in the DNA^{9C} structure is "smeared" to the point that the density overlaps with the secondary layer waters, while the electron density of the water in DNA^{9T} structure discrete and well-determined. Although this is a subtle change in the structure, the difference in the electron density of this primary water likely represents a real difference in the stability of the spine of hydration in the non-contacted region between the P22R NTD- DNA^{9C} structure and the P22R NTD- DNA^{9T} structure. This observation is perhaps the most significant difference between the DNA^{9C} and the DNA^{9T} complexes. The stability of the B' state is dependent on the ability of double stranded DNA to establish and maintain a spine of hydration. The less stable of the spine of hydration in the non-contacted region between the DNA^{9C} due to the TpA→CpG

mutation at the central step would give rise to a less stable B' state. The higher energy B' state of the DNA^{9C} non-contacted region would in turn give rise to a higher energy complex. The prediction of a higher energy state complex due to the higher energy state of the DNA conformation would explain the 10 fold lower affinity of the repressor for an operator with a TpA→CpG mutation at the central non-contacted step of the sequence.

The functional groups on the floor of the minor groove of AT rich DNA are ideal for the coordination of ordered water molecules and monovalent cations due to presence of hydrogen bond acceptor and absence of hydrogen bond donors. The presence of hydrogen bond donors and absence of hydrogen bond acceptors gives rise to a high polarity and high electronegative potential. The presence of the exo-cyclic amines of the guanines at position 9L and 9R of the DNA^{9C} structure would introduce a hydrogen bond donor into the minor groove, changing the polarity and electrostatic potential of the groove.

Furthermore, the hydrogen bonding in the minor groove of A•T and T•A base pairs allows for the propeller twisting and base pair opening necessary for optimal coordination of primary solvent molecules without straining base pair hydrogen bonds. The O2-NH2 hydrogen bond is likely strained by the propeller twisting at the CpG step that is necessary to coordinate the primary water of the spine of hydration in the non-contacted region. The presence of the minor groove amines of the guanines and the addition of minor groove hydrogen bond by the C•G and G•C base pairs likely disrupt or reduce the stability of the spine of hydration. A disruption in the spine of hydration would results in a less favorable B' state, and thus a less stable complex between the P22 c2 repressor and its operator.

The Tunnel

The tunnel observed in the P22R NTD-operator complexes is formed by the dimer interface and the minor groove of the non-contacted region. This structural device is likely a key element in the indirect readout mechanism by the P22 c2 repressor. First, the tunnel insures the proper conformation of the sugar-phosphate backbone in the non-contacted region. The backbone nitrogen and the δN of Asn49 on both the L and R monomers chelate the O2 of the phosphate of residue 13 and 33 of the DNA, respectively. The interaction of the protein residues of "ceiling" of the tunnel with phosphate oxygens of the "walls" modulates the proper orientation of the repressor with respect to the operator. Secondly, The δNH_2 of Asn46 located in the "ceiling" of the tunnel hydrogen bonds with the tertiary layer waters in the spine of the non-contacted region. The coordination of the waters by the repressor on the ceiling of the tunnel may enhance the stability of the spine so that less than optimal interactions of solvent on the floor of the tunnel are tolerated. Finally, the tunnel motif separates the minor groove of the non-contacted region from the entrance and exit of solvent molecules from the top of the groove, and forces solvent exchange to occur at the entrance and exit of the tunnel. The reduced solvent exchange in the tunnel may result in increased localization of water and cations in the minor groove of the non-contacted bases. The increased localization of solvent would help stabilize the B' state by helping maintain the spine of hydration.

Conclusion

For P22 c2 repressor to be properly oriented relative to a DNA operator, the groove of the operator must be narrow in the center of the operator. The narrow minor groove is characteristic of the B'-state. The B' state is thermodynamically linked to the

presence of a spine of hydration and the reduced electrostatic repulsion of the negative cross-strand phosphate groups. The transition from B→B'state is sequence dependent. In free DNA, A-tracts have high propensities to conform to the B' state(*38, 85, 117*). This propensity is partly due to the proper positioning chelating O2 atoms of thymine to coordinate water and cations.

The structure of the DNA^{9T} complex shows the presence of a highly ordered spine of hydration. The primary layer of the spine is coordinated by N3's of cross-strand adenines at TpA step. While the N3-W-N3 coordination is likely less preferred than that of coordinate site involving the O2 of a thymine (*40, 43, 59*), The propensity of propeller twisting by A•T and T•A base pairs, the stability added by the tertiary water stabilization by the ceiling of the tunnel, and interactions with the phosphate backbone and the repressor may minimize any loss of stability in the primary water coordination.

We have shown that the binding of P22R NTD induces a narrow minor groove as well as a spine of hydration despite the presence of a CpG step. The CpG step does however result in more disorder in the spine , particularly in the floor of the minor groove. The change in the spine of hydration is the only observable difference between the x-ray crystal structures of the DNA^{9C} and the DNA^{9T} complexes. We propose that the reduced stability of the spine results in a higher energy cost for the narrow minor groove. The higher energy cost for a narrow minor groove results in a higher energy state of the complex. These findings are agreement with the solution studies that the P22 c2 repressor binds a DNA operator with a CpG at the central non-contacted step with lower affinity than that with a TpA at the central step.

CHAPTER 4

ELECTROSTATIC FORCES IN THE P22 C2 REPRESSOR-DNA COMPLEXES

Introduction

The DNA base pairs at the center of the P22 c2 operator are not contacted by the repressor, yet changes in base identity at this central region results affects the affinity of the repressor for the operator (*23, 29*). This mode of discrimination without direct or water mediated contacts between the protein and DNA bases has been termed 'indirect readout' of DNA sequence (*27*). 'Direct readout' is defined as the sensing of base pair identity by direct hydrogen bonding and van der Waals interactions with functional groups on the DNA base pairs. We have proposed that the mechanism of indirect readout by the P22 c2 repressor involves the induction of the B' state in the central region of the operator [citation]. This model is based on the crystal structure of the P22R NTD (P22 c2 repressor N-terminal domain residues 1-68) has been solved to 1.6 Å resolution with the DNA operator DNA^{9T}.

The B'-state of DNA (*118*) differs from the standard B-DNA by (i) a narrow minor groove (*81, 82*), (ii) a spine of hydration (*37, 39*) and higher-order hydration within the minor groove (*58, 83*), (iii) a preference for sequence elements of at least three contiguous ApA or ApT steps (5' to 3', TpA steps not favored, G-C base pairs not allowed), (iv) a propensity to bend DNA when appropriately phased (*78-80*), (v) negative propeller twisting of base pairs (*85, 86*), (vi) unusual rigidity (*88, 119*), and (vii) a 'pre-melting' thermal transition that is distinct from duplex melting (*89-91*), and (viii) monovalent ions within the minor groove (*38, 53, 55, 56, 84*). The B' helical axis is

thought to be linear (*92, 93*); axial bends arise at the junctions between B' and flanking B-DNA (*36*). B' hydration, especially within the minor groove, appears to be cooperative, with a many interdependent solvent interactions. B' DNA and B-DNA along with other helical forms demonstrate that DNA conformation is polymorphic (reviewed by Hud and Plavac (*36*)) and Rich (*45*)).

Here we describe structures obtained from crystals grown in the presence of heavy monovalent cation substitutes. We have determined structures of P22R NTD in complexes two different operators, DNA^{9T} and DNA^{9C} and have mapped out electrostatic forces and positions of monovalent cations in these complexes. The two operators differ at the central base step, which is mutated from TpA→CpG for DNA^{9T}→DNA^{9C} (Refer to Figure). Neither DNA^{9T} nor DNA^{9C} contain an A-tract in the non-contacted region.

Monovalent cations are not observed in either of the previous structures of the P22R NTD-DNA complexes. Nor are monovalent cations observed in other helix-turn-helix DNA complexes or to our knowledge, in DNA-protein complexes in general. We have addressed the question of whether the absence of monovalent cations in previous DNA-protein complexes results from restrictions on ability of standard x-ray crystallographic techniques to distinguish Na^+, K^+ and NH_4^+ ions from water molecules. In this work, we have determined the position of localized monovalent cations in the P22R NTD-DNA^{9T} and the P22R NTD-DNA^{9C} complexes by co-crystallizing the complexes with heavy atom K^+ mimics. To determine the positions of monovalent cations we crystallized the P22R NTD-DNA^{9T} complex in the presence of thallium (Tl^+) and the P22R NTD^{9C} complex in the presence of rubidium (Rb^+) (*54, 84, 120*). Both Tl^+ and Rb^+ have been shown to mimic K^+ in interactions with DNA [refs], RNA [refs] and

proteins (*121-124*). The advantage of Tl^+ and Rb^+ in crystallographic studies, is that they are directly observable, without resorting to analysis of subtle differences in coordination chemistry, by ther strong scattering of x-rays and their anomalous scattering.

Materials and Methods

Construction of plasmid expressing P22R NTD

A DNA fragment encoding the NTD fragment of P22 repressor was created using PCR to insert a stop codon at position 69 of the intact P22 repressor gene. We amplified a 450 bp fragment from pTP125 (*64*) using primers obtained from the CAMBI Nucleic Acid Facility (University at Buffalo). The sequences of the primers were: 5'GTTTTTTGCGTCGACATCATA 3' and 5' CGTTTGTCTGAC GTCAATCTCCTTTC 3'. This fragment, containing the tac promoter (*65*) fused to the P22 repressor gene fragment, was cleaved with Aat II and Sal I, gel purified, ligated in pUC18 cut with the same enzymes and transformed into E. coli strain XA90 (*66*) . The resulting plasmid expresses a protein comprised of the first 68 amino acids of P22 repressor.

Purification of P22R NTD

E. coli XA90 cells (*66*) bearing P22R NTD expression plasmid were grown with aeration at 37°C in 3 liters of Luria broth supplemented with 0.1 mg/mL ampicillin to an A_{600} of approximately 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mM, and growth at 37°C was continued for 4 hours. The cells were harvested by centrifugation and resuspended in 20 mL lysis buffer (200 mM NaCl, 50 mM EDTA; 100 mM Tris pH 7.4) with 0.5 mL of 20mg/mL phenyl methyl sulfonyl fluoride (PMSF). The

cells were lysed in a French pressure cell (18,000 lb/in^2) at 4^{o}C. The lysate was diluted with 100 mL of cold lysis buffer. Cellular debris was precipitated by centrifugation at 4000 x g. Nucleic acids were precipitated by dropwise addition of polyethyleneimine to a final concentration of 0.6% at 4^{o}C and mixed for 15 minutes, followed by centrifugation at 10,000 x g. Soluble proteins were precipitated with 40% (w/v) ammonium sulfate and dialyzed at 4^{o}C against three 1 L changes of standard buffer (SPB-50 mM $NaPO_4$ pH 6.8 10 mM EDTA) supplemented with 50 mM NaCl. The dialyzate was loaded onto three Econo-Pac carboxymethyl (CM) cartridges (Bio-Rad, Hercules, CA) and eluted with a linear salt gradient from 0.1 to 1 M NaCl in SPB over 70 minutes. Fractions containing repressor were pooled and precipitated with 40% ammonium sulfate. Following precipitation, the protein-containing pellet was dissolved in SPB+100 mM NaCl and loaded onto a 1 cm X 50 cm Sephacryl S100 (Amersham Biosciences) size exclusion column. The column was developed with the same buffer at a flow rate of 0.3 mL/min. Fractions containing P22R NTD were pooled and the protein was precipitated with 40% ammonium sulfate. For Tl^+ derivative crystal, the dissolved pellet was dialyzed against 50 mM TrisAcetate (7.5), 100mM TlAc, and 10% glycerol. For Rb^+ derivative crystal, the dissolved pellet was dialyzed against 50 mM Tris-HCl (7.5), 100mM RbCl, and 10% glycerol The protein was flash frozen in liquid nitrogen and stored at -70°C.

DNA purification

Both DNA^{9T} and DNA^{9C} oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA) as an ammonium salt from reverse HPLC purification. To form the double-stranded protein binding site, equimolar amounts of the complementary

strands were mixed in water, heated to 85°C for 60 s and slow-cooled over four hours to anneal.

Co-crystallization of P22R NTD with DNA^{9T}/Tl^{+}

Crystals were grown at 4°C by the hanging drop vapor diffusion method. The initial crystallization solution contained 0.42 mM P22R NTD, 0.42 mM duplex d($^{5'}$CATTTAAGATATCTTAAATG$^{3'}$) -d($^{5'}$CATTTAAGATATCTTAAATG$^{3'}$), 45 mM Tris·Ac (pH 7.8), 20 mM TlAc, 2.0% glycerol, 14% PEG 400, 2.3mM $MgAc_2$ and 0.9% MPD in a volume of 5.7 µl. The crystallization solution was equilibrated against a reservoir of 100 mM Tris·Ac (pH 7.8), 25% PEG 400, 5 mM $MgAc_2$ and 2% MPD. Crystals grew to a size of approximately 0.5x0.3x0.3 mm^3 within one day. A single crystal was collected from the crystallization solution in a nylon loop and flash frozen in liquid nitrogen.

Co-crystallization of P22R NTD with DNA^{9C}/Rb^{+}

Crystals were grown at 4°C by the hanging drop vapor diffusion method. The initial crystallization solution contained 0.42 mM P22R NTD, 0.42 mM duplex d($^{5'}$CATTTAAGACGTCTTAAATG$^{3'}$) -d($^{5'}$CATTTAAGACGTCTTAAATG$^{3'}$), 45 mM Tris·HCl (pH 7.8), 20 mM RbCl, 2.0% glycerol, 14% PEG 400, 2.3mM $MgCl_2$ and 0.9% MPD in a volume of 5.3 µl. The crystallization solution was equilibrated against a reservoir of 100 mM Tris·HCl (pH 7.8), 25% PEG 400, 5 mM $MgCl_2$ and 2% MPD. Crystals grew to a size of approximately 0.5x0.3x0.3 mm^3 within one day. A single crystal was collected from the crystallization solution in a nylon loop and flash frozen in liquid nitrogen.

Data Collection of P22R NTD-DNA^{9T}/Tl^{+}

Diffraction data were collected on SER-CAT beamline 22 BM at the Advanced Photon Source, Argonne National Laboratory with a MAR225 CCD detector using radiation of 0.97784 Å wavelength and a crystal-to detector distance of 150 mm. Images were collected with a 3 s exposure time and a 1° oscillation. The crystal was maintained at 110 °K. A total of 452,190 reflections were indexed and integrated with DENZO (*67*) and reduced to 30,146 unique reflections in space group $P4_3$ in the resolution range of 30-1.9 Å. Data were scaled with the program SCALEPACK(*68*).

Data Collection of P22R NTD-DNA^{9T}/Tl^{+}

Diffraction data were collected on SER-CAT beamline 22 BM at the Advanced Photon Source, Argonne National Laboratory with a MAR225 CCD detector using radiation of 0.81398 Å wavelength and a crystal-to detector distance of 200 mm. Images were collected with a 5 s exposure time and a 1° oscillation. The crystal was maintained at 110 °K. A total of reflections were indexed and integrated with DENZO (*67*) and reduced to 22,529 unique reflections in space group $P4_3$ in the resolution range of 41-2.1 Å. Data were scaled with the program SCALEPACK(*68*) with Bijvoets separated.

Refinement

Initial phases were determined by molecular replacement with the program Molrep of the CCP4 suite using data from 15 Å to 3Å, and the P22NTD-DNA^{9T} complex with waters removed was used a search model. The data was extended to 1.9 Å where R_{sym} were reasonable, and the model was refined in Refmac5. Further model building was performed by fitting proper protein sidechain rotomers in COOT. Water molecules were added iteratively to the model initially using ArpWarp and by inspection of $|2F_o\text{-}F_c|$

and $|F_o-F_c|$ maps in later stages, followed by cycles of refinement in Refmac5, resulting in a R-factor of 18.4% and a R_{free} of 20.8%. Data collection and refinement statistics are given in Table 4.1.

Initial phases were determined by molecular replacement with the program Molrep of the CCP4 suite (*111*) using data from 15 Å to 3Å, and the P22NTD-DNA^{9C} complex with waters removed was used a search model. The data was extended to 2.1 Å where R_{sym} were reasonable, and the model was refined in Refmac5. Further model building was performed by fitting proper protein sidechain rotomers in COOT. Water molecules were added iteratively to the model initially using ArpWarp and by inspection of $|2F_o-F_c|$ and $|F_o-F_c|$ maps in later stages, followed by cycles of refinement in Refmac5, resulting in a R-factor of 18.3% and a R_{free} of 22.4%. Data collection and refinement statistics are given in Table 4.1.

Anomalous (F_+-F_-) and Isomorphous (F_{Tl+}-F_{Na+}) Difference Maps

The F_+-F_- Fourier maps were calculated for DNA^{9T}/Tl^+ and DNA^{9C}/Rb^+ complexes using the FFT from the CCP4 Suite (*111*). Phases for the anomalous maps were calculated from the respective models with solvent molecules removed, using SFall of the CCP4 suite. Isomorphous difference maps were calculated using the data from the P22R NTD-DNA^{9T} complex (chapter 2) as the native derivative (Na^+). Data sets were merged using CAD. P22R NTD-DNA^{9T}/Tl^+ data were scaled to the P22R NTD-DNA^{9T} native data using Scaleit, both from the CCP4 suite. Phases for the F_{Tl+}-F_{Na+} were calculated from the P22R NTD-DNA^{9T}/Tl^+ model with all solvent sites treated as water molecules.

Table 4.1. Data Collection and Refinement Statistics

	P22R NTD-DNA^{9T}/Tl^{+}	P22R NTD-DNA^{9C}/Rb^{+}
X-ray source	SER-CAT 22-BM APS	SER-CAT 22-BM APS
Wavelength (Å)	0.97784	0.81398
Detector	MAR CCD 225	MAR CCD 225
Resolution Range (Å)	30-1.90	41-2.10
Mosaic spread (°)	~0.5	~0.7
Space Group	$P4_3$	$P4_3$
Unit-cell parameters (Å)		
A	64.249	64.007
B	64.249	64.007
C	100.820	101.623
Total Reflections	452,190	347,031
Unique Reflections	30,146	22,529
Multiplicity	13.4 (11.3)[a]	12.5 (10.4)[a]
Completeness (%)	98.5 (97.7)	99.3 (99.6)
Highest Resolution Shell (Å)	1.95-1.90	2.15-2.10
$I/\sigma(I)$	44.3 (9.29)	33.2 (4.5)
Rsym (%)	7.3 (48.7)	14.2 (63.6)
Reflections (working/test set)	30,146/1591	22,529/1214
Number of non-H protein atoms	2 x 514	2 x 514
Number of non-H DNA atoms	814	814
Number of water molecules	304	296
Number of cations	3/10[b]	2
R-crystal	18.4%	18.0%
R-free	20.8%	22.5%
Average B-factor	25.06	26.55
RMSD from ideal		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	1.7	2.2

[a]The number in parenthesis indicates the value for the highest resolution shell. [b]Number of cations included in final model/number of cations indicated by both anomalous maps and isomorphous difference maps

Electrostatic Calculations

ESP calculation was performed on the P22R NTD alone by solving the non-linear Poisson-Boltzmann equation using the Adaptive Poisson Boltzmann Solver (APBS) (*125*) with the Pymol interface. Calculation were ran at 150 mM NaCl with a probe size of 1.4 Å. The van der Waals radius and partial charges of the atoms of the protein and the DNA were calculated using Amber force fields as implemented in PDB2PQR (*126*).

Results

Cation localization of the P22R NTD-DNA complexes

Anomalous difference F_+-F_- density of P22R NTD-DNA^{9T} Tl^+ derivative

The anomalous difference map, synthesized using F_+-F_- as Fourier coefficients, contains three peaks that indicate sites of localized anomalous scatters (Tl^+ ions, Figure 4.1). These three sites overlap solvent peaks of the $2F_o$-F_c Fourier maps, suggesting that these water sites in the previous structure may be in fact be partially occupied by sodium or ammonium ions. Anomalous peaks are observed that correspond to phosphorous atoms of the DNA backbone.

All of the solvent anomalous peaks are located in the major groove. The site of the first Tl^+ peak in the major groove (site Tl^{+MG1}, 3.5 σ) is coordinated by the O6 and N7 of G(28) at position 7L of the operator and is modeled at 20 % occupancy. The Tl^{+MG1} site is not symmetric about the operator, in that the equivalent site ($Tl^{+MG1'}$) in proximity of the O6 and N7 of G(8) at position 7R is occupied by a cationic sidechain of the protein, not a Tl^+ ion. The O6 and N7 of G(8) do form hydrogen bonds with the positively charged amine of a crystallographic symmetry related lysine at position 7R.

The lysine interactions at $Tl^{+MG1'}$ provides a electrostatic environment that is similar to that observed for Tl^{+MG1}.

The two most intense anomalous highest peaks are located at the DNA junction of symmetry related CpG step. Site Tl^{+MG2} (16σ), with an estimated occupancy of 25%, is coordinated by the O6 and N7 of G(40) of the terminal base pair. $Tl^{+MG2'}$ (26 σ) is modeled at 35% occupancy, and is coordinated by the O6 and N7 of G(20) of terminal base pair at the opposite end of the operator from site Tl^{+MG2}. Crystal packing aligns the helical axis of symmetry related DNA duplexes such that the C(1)•G(40) base pair stacks onto the C(21)•G(20) base pair. Tl^{+MG2} and $Tl^{+MG2'}$ are at opposite ends of the DNA operator; symmetry brings them within close proximity of each other. There is significant overlap of the two anomalous peaks. So, while the sites have been modeled as two separate Tl^+ cations whose combined occupancy is ~70%, it may be likely that the peak represents a single cation whose position is distributed between the two sites of Tl^{+MG2} and $Tl^{+MG2'}$ (Figure 4.2).

F_{Tl}^{+}-F_{Na}^{+} Isomorphous difference density of P22R NTD-DNA^{9T}

To confirm the peaks of the anomalous map and to attempt to determine additional heavy atom sites (cation binding sites), an isomorphous difference map was calculated using the diffraction data from the P22R NTD-DNA^{9T}/Na^+ derivative (see chapter 2) and the P22R NTD-DNA^{9T}/Tl^+ derivative. The reflection data merged with an R_{iso} of 28.5% averaged over all shells (*127*). The isomorphism between the two structures was further confirmed by an RMSD of 0.41 Å for all backbone atoms of the DNA and protein. Differences in unit cell axis are Δa= 0.2% ,Δb= 0.2%, and Δc= 0.8%. By comparison with other successful efforts to identify Tl^+ sites with isomorphous

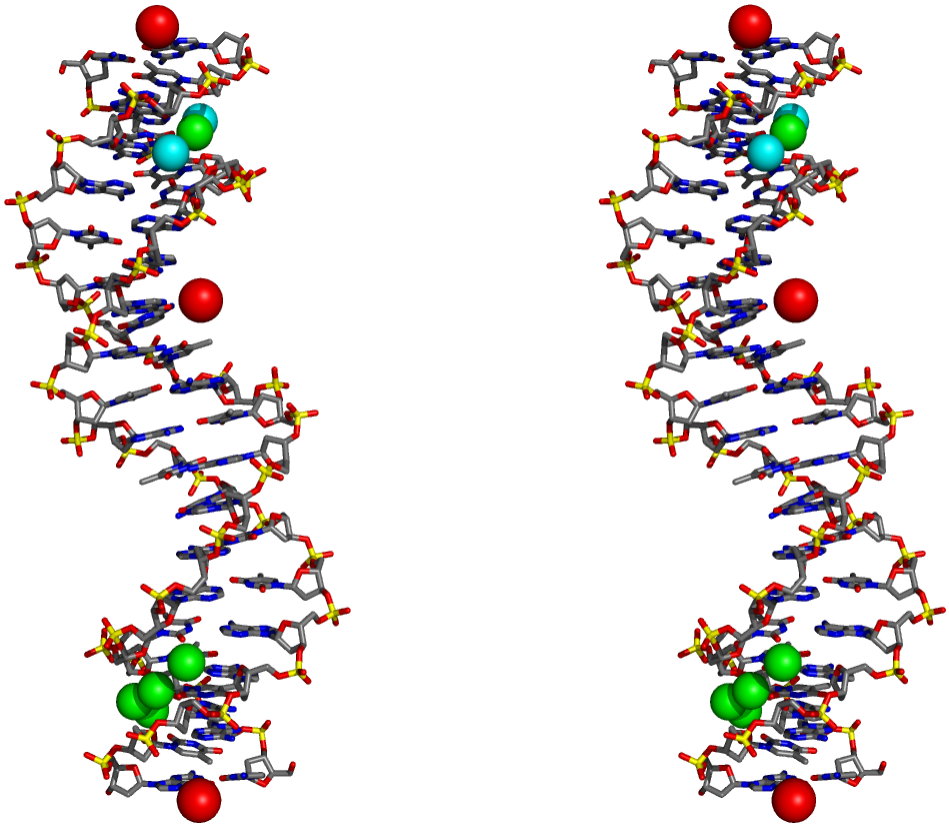

Figure 4.1. Monovalent cation binding sites identified by anomalous peaks and F_{Tl+}-F_{Na+} are shown in red spheres. Sites identified by isomorphous peaks only with F_{Tl+}-F_{Na+} peaks $\geq 3.5\sigma$ are shown as green spheres and F_{Tl+}-F_{Na+} peaks $\geq 3.0\sigma$ are cyan spheres.

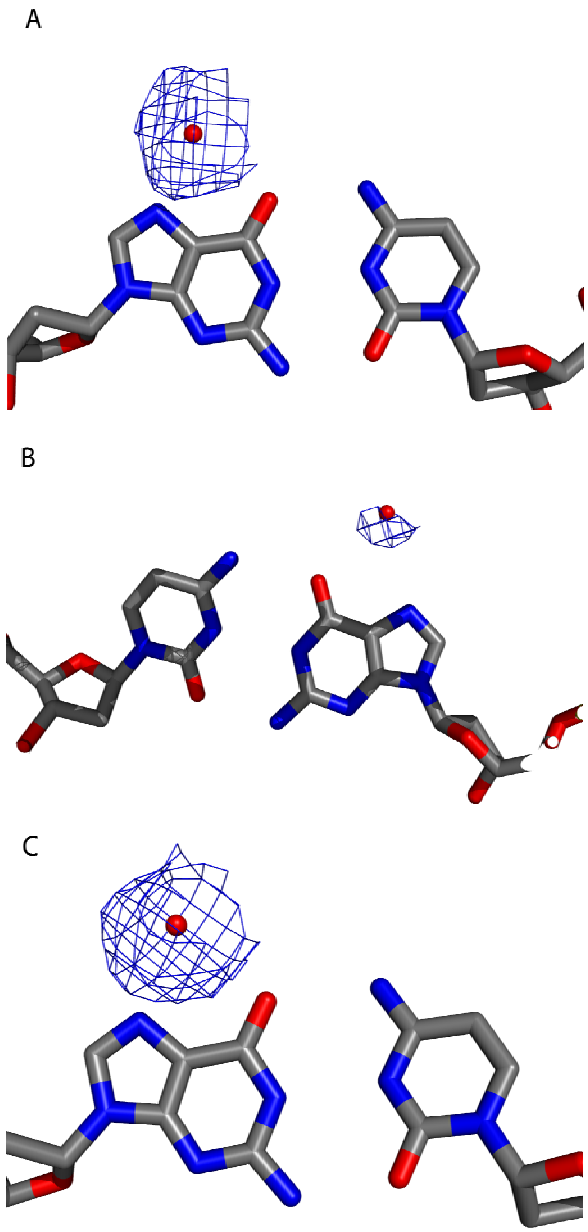

Figure 4.2. Anomalous difference peaks calculated from a F_+-F_- Fourier map contoured at 3σ. Panel A is the $Tl^{+MG2'}$ site coordinated on the R-terminal of the operator by O6 and N7 of G(40). Panel B is the Tl^{+MG1} site coordinated at position 7L of the operator by O6 and N7 of G(28). Panel C is the Tl^{+MG2} site coordinated on the L-terminal position of the operator by O6 and N7 of G(20).

difference (*54, 128*), the statistics indicate that the P22R NTD complexes used here are very highly isomorphous.

We have identified 10 isomorphous difference peaks that correspond to solvent sites in the 2Fo-Fc maps of P22R NTD-DNA^{9T} (Figure 4.1). Isomorphous difference methods are often less accurate than anomalous scattering methods except for highly isomorphous structures. Here we consider the observed isomorphous difference peaks to highly reliable because (i) the non-crystallographic symmetry is reasonably preserved in that most isomorphous difference peaks are observed in the same places at both ends of the complex, making the same interactions, and with the same relative intensities, (ii) the observed isomorphous difference peaks are located in the positions predicted from a previous structure of a DNA oligomer crystallized in the presence of Tl^+ (*54, 120*)

Seven of the isomorphous difference peaks are located in the minor groove of the A-tracts on each end of the operator. Six of these interact directly with functional groups on the floor of the minor groove, at sites were previously identified as primary waters of the spine of hydration in the P22R NTD-DNA^{9T}/Na^+ complex. In addition to the functional groups on the floor of the minor groove, each site is coordinated by two O4' of the sugars displaced by one residue in the 5' direction, and by secondary water molecules of the spine of hydration. The positions and coordination of the five reasonable sites identified here are equivalent to the positions and coordination of monovalent cations previously observed in the A-tracts of unbound DNA crystal and NMR structures (*54-56, 84, 120, 129*)

Tl^{+mg1} (3.5σ) is located between positions 1R and 2R, and is coordinated by the O2 of T(39) and the O2 T(3) (Figure 4.3). The equivalent site at the other end of the

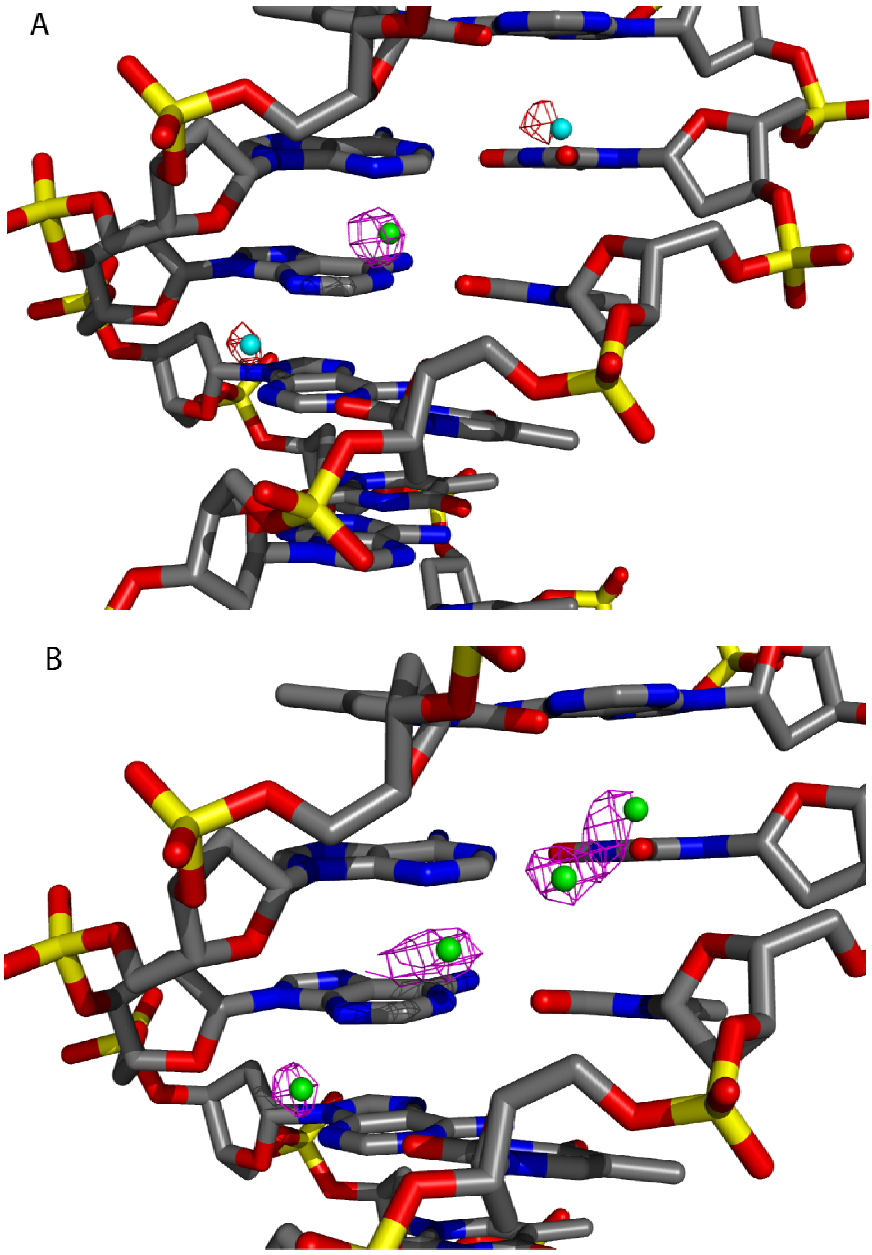

Figure 4.3. Minor groove cation sites identified from F_{Tl+}-F_{Na+} peaks with magenta density contoured at 3.0σ and F_{Tl+}-F_{Na+} with red density contoured at 2.5σ. Panel A shows the sites from the P22R NTD-DNA^{9T}/Tl^{+} derivative on the L-half of the operator and Panel B shows the sites from the R-half of the operator.

complex, $Tl^{+mg1'}$ (3.0σ) is located between positions 1L and 2L and is coordinated by O2 of T(19) and O2 of T(23).

Tl^{+mg2} (3.5σ) is coordinated by the N3 of A(38) (2R) and O2 of T(4) (3R). $Tl^{+mg2'}$ (5σ) is coordinated by the N3 of A(18) (2L) and O2 of T(24) (3L) and the O4' atoms of the deoxyriboses of T(19) and T(25).

Tl^{+mg3} (3.5σ) is coordinated by N3 of A(37) (3R) and O2 of T(6) (4R). $Tl^{+mg3'}$ is coordinated by the N3 of A(17) (3L) and O2 of T(25) (4L). Tl^{+mg4} (3.5σ) is located in the secondary layer of the spine of hydration, and is perched above the primary sites of Tl^{+mg1} and Tl^{+mg2}. At a contour of 3.0σ the difference peak of Tl^{+mg4} is not fully distinct from the peak of Tl^{+mg1}.

Anomalous difference F_+-F_- density of P22R NTD-DNA^{9C} Rb^+ derivative

The anomalous difference map of the P22R NTD-DNA^{9C}/Rb^+ derivative reveals two peaks located in the major groove, Rb^{+MG1} and $Rb^{+MG1'}$, which are coordinated by guanines at the CG step created by the junction of two symmetry-related DNA helices. The two Rb^+ anomalous peaks are analogous to the two highest Tl^+ anomalous peaks of the DNA^{9T} complex (see above). The smaller peak (3.9σ) is equivalent in position to highest peak $Tl^{+2MG'}$ (26σ) for the DNA^{9T}. The Rb^+ occupancy of this site is estimated at 30%.

$Rb^{+MG1'}$ is coordinated by the O6 and N7 of G(20) of the terminal base pair. The highest peak, Rb^{+MG2}, (5σ) corresponds Tl^{+MG2}. Rb^{+MG1} is modeled at 30% occupancy and is coordinated by the O6 and N7 of G(40) of terminal base pair at the opposite end of the operator (Figure 4.4).

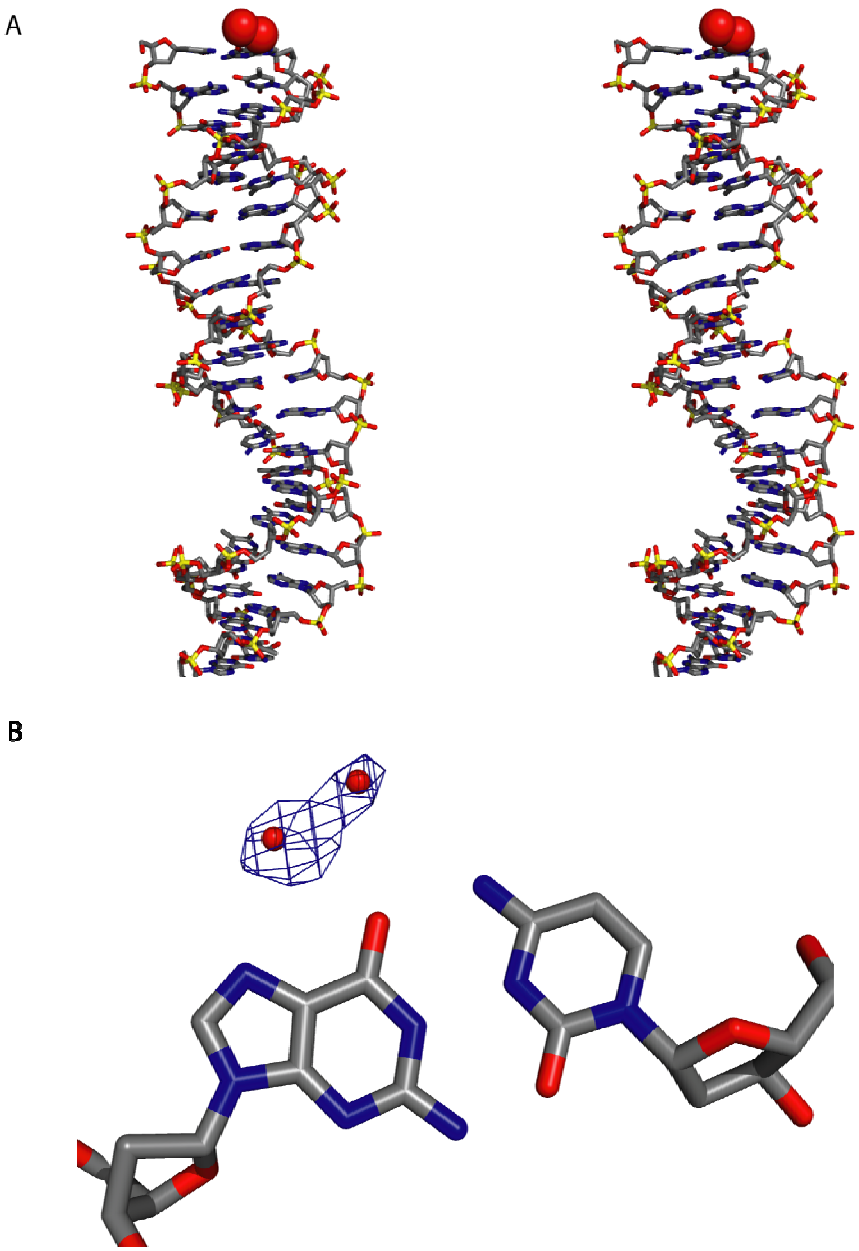

Figure 4.4. Rubidium sites identified by anomalous peaks form F_+-F_- Fourier map of P22R NTD-DNA^{9C}/Rb^+ derivative. Panel A shows the stereo view of the cation sites relative to DNA^{9C} operator. Panel B shows density surrounding the sites contoured at 3.0σ.

As with the DNA^{9T}/Tl^{+} derivative, the crystal packing aligns the helical axis of symmetry related DNA duplexes in the DNA^{9C}/Rb^{+} derivative, resulting in the stacking of the C(1)•G(40) base pair onto the C(21)•G(20) base pair. Thus, the crystal packing, again, brings sites $1Rb^{+}$ and $2Rb^{+}$ within close proximity, and there is significant overlap of the two anomalous peaks. Therefore, it is likely that the peaks of Rb^{+1} and $Rb^{+1'}$ represent a single cation site whose position is distributed between the two sites.

Cation Localization in the Minor Groove of the NC region

The structures of the P22R NTD-DNA^{9T}/Tl^{+} and the P22R NTD-DNA^{9C}/Rb^{+} derivatives give no indication of localized monovalent cations in the minor groove of the non-contacted region. A close examination of the anomalous maps and isomorphous difference maps was performed, and no peaks were observed that could be distinguished from the noise. None of cation localization sites are located in close proximity to the non-contacted region of the P22R NTD operators.

Electrostatic Surface Potential (ESP)

ESP of the P22R NTD

ESP calculations reveal that the most positive region of the repressor is at ~8 kT/e- and occurs at the outer edges of the protein that contact the phosphates at each end of the repressor. The most negative part of the repressor is at the center of the repressor slightly above the DNA face of the protein at the entrance to the tunnel region. This electronegative region corresponds to the position Glu44 and Glu48 and comes in close proximity the phosphates of the DNA in the complex (Figure 4.5).

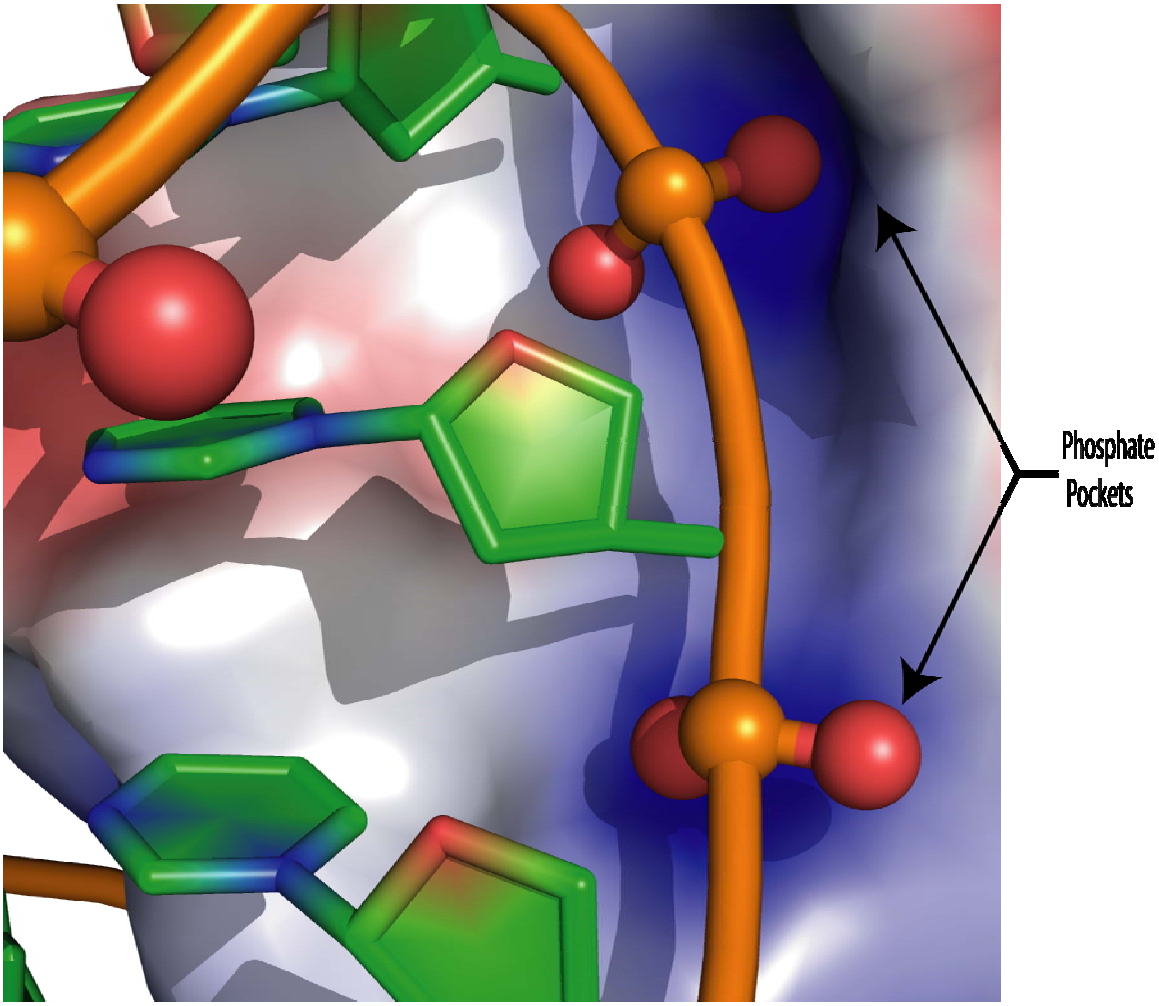

Figure 4.5. The Phosphate Pockets. Graphical representation of ESP calculated at the solvent accessible surface area of the P22R NTD alone. P22R NTD shown as surface colored red, -3 kT/e^-, to blue +3 kT/e^-. DNA backbone shown in tubes with phosphate oxygens shown in red spheres. DNA bases shown as cartoon. Intense blue represents the phosphate pockets on the surface of the P22R NTD.

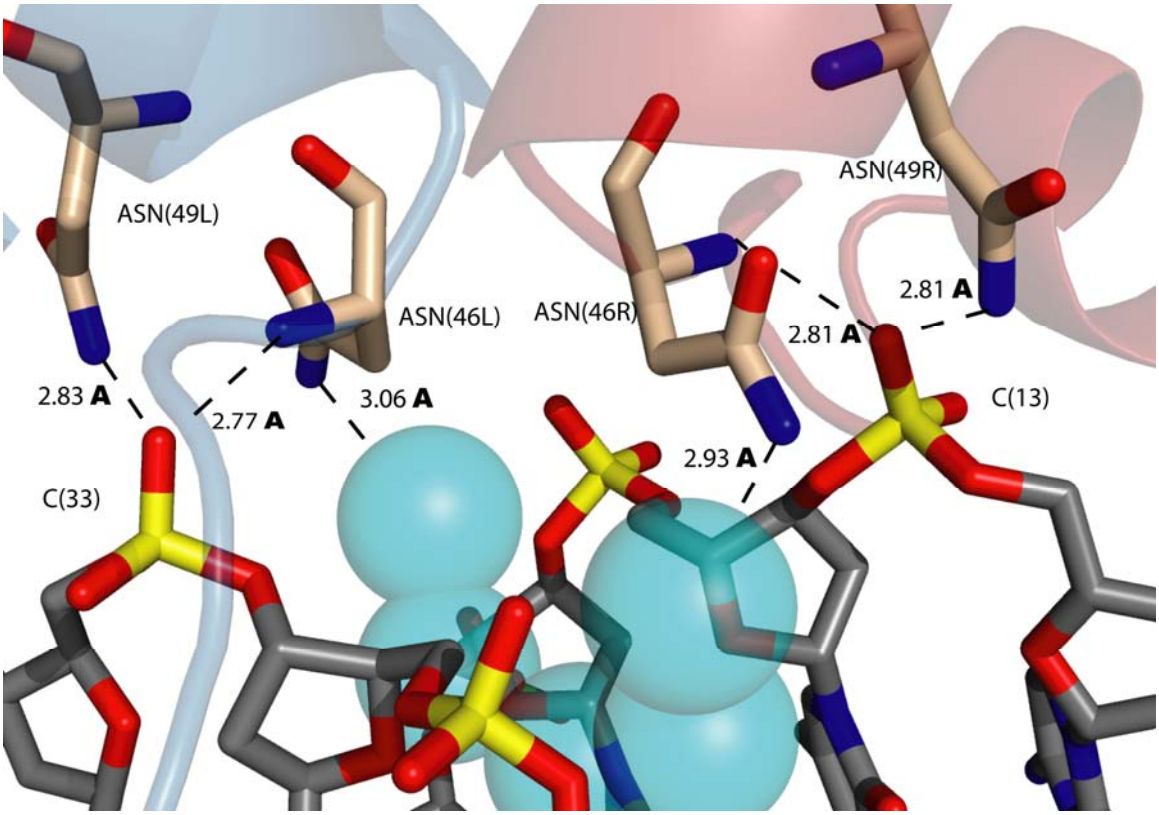

Figure 4.6. Protein interactions with the DNA backbone and waters of the spine of hydration in the minor groove of the non-contacted region. Interacting protein residues and DNA backbone is shown in stick. The remaining protein is shown in ribbons. Water and ribbons of the protein are set partially transparent for clarity.

The DNA face of the repressor has a positive ESP that runs approximately diagonal across the face of the repressor. This strip of positive potential contains four discrete regions of high positive ESP. These four regions of positive regions create "phosphate pockets" for the phosphates in the deoxyribose-phosphate backbone. The two most central "phosphate pockets" align with the phosphates in the non-contacted region where the groove is narrowest. This alignment creates a complementary electrostatic surface between the repressor and DNA that is in the B' state.

ESP of the P22R NTD-DNA^{9T} complex

ESP calculation of the P22R NTD-DNA^{9T} complex show that the most negative potential of the complex in the tunnel region. The floor of the minor groove in the non-contacted region has a more negative potential than the floor of the A-tracts on each end of the DNA, -16 kT/e$^-$ for the non-contacted region compared to -11 kT/e$^-$ for the A-tracts. The minimum potential in the induced narrow minor groove calculated for the DNA alone is the same as that for the A-tracts on either end of the DNA, -11 kT/e$^-$. A comparison of the ESP of the DNA alone versus ESP of the complex indicates that the more negative potential of the narrow minor groove in the tunnel is a result of P22R NTD binding (Figure 4.7).

Discussion

A previous x-ray structure of a P22R NTD-DNA complex(*115*) indicates that binding of P22R induces the B' state in the NC region, with a narrow minor groove and a highly ordered spine of hydration. Those observations are confirmed here, for additional NC sequences, including those with GC base pairs. Combined, these observations support a model in which the NC sequence is read at a distance by its propensity to adopt the A'

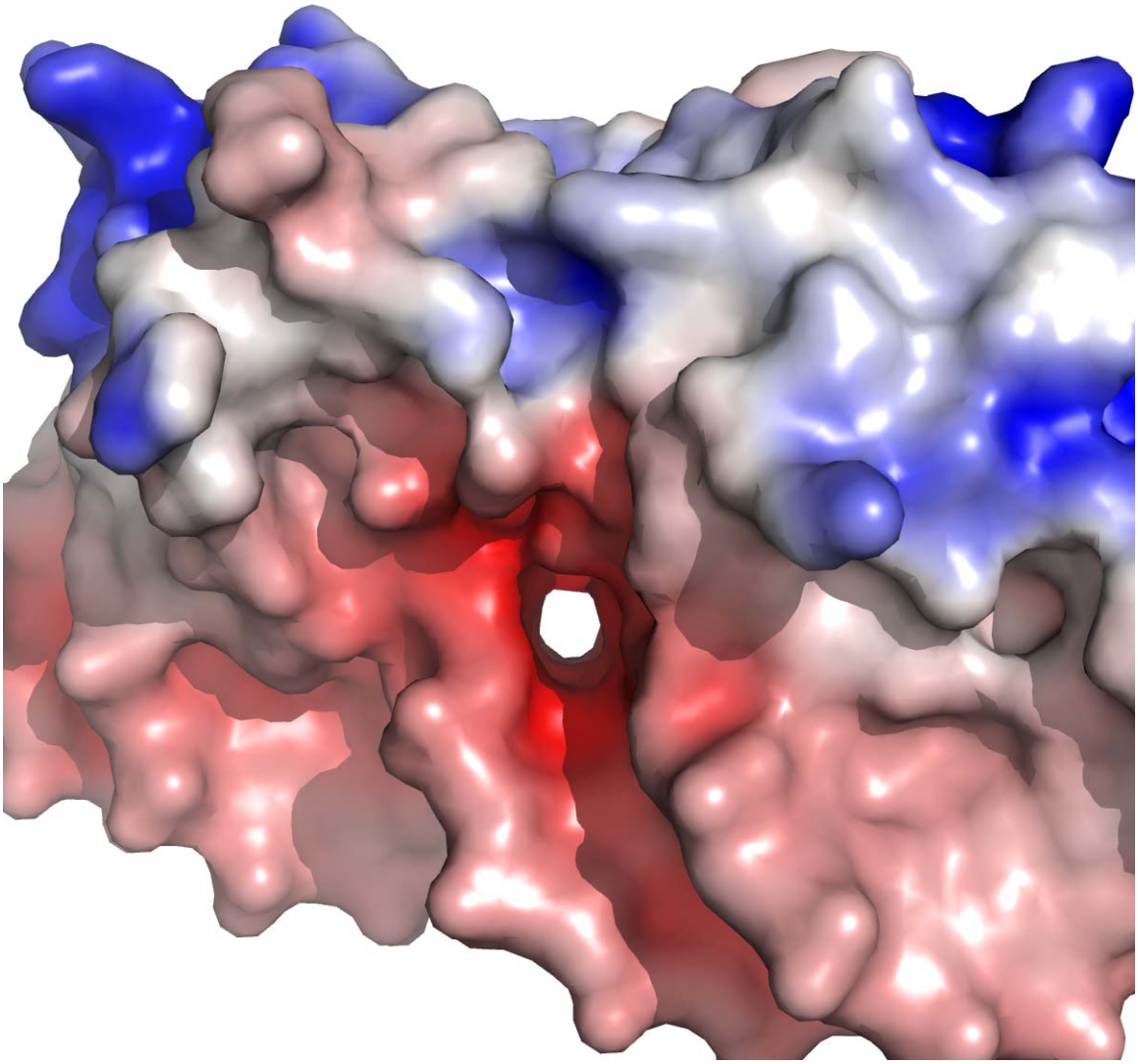

Figure 4.7. ESP of the P22R NTD-DNA^{9T}. Surface representation of ESP calculated at the solvent accessible surface area of the P22R NTD-DNA complex. View is looking along the minor groove in tunnel region of the P22R NTD-DNA^{9T}. Surface colored red, -10 kT/e$^-$, to blue +3 kT/e$^-$.

state. Sequences that readily adopt the A' state bind with high affinity. Sequence that resist the A' state bind with lower affinity. This model is generally consistent the Koudelka's biochemical data.

Complimentary ESP of the P22R NTD and the B' State

Electrostatic surface potential (ESP) calculations have previously shown that the minor groove becomes increasingly electronegative as is narrows (*36*). Here ESP calculations indicate that both narrowing and P22R binding contribute to the electronegative potential of the NC minor groove. Thus the electronegative potential in induced B' state of the P22R NTD-DNA complexes (i.e., in the NC region) is greater than naturally-occuring B' state associated with the A-tracts.

The most positive potential on the DNA face of the P22R NTD is aligned with phosphates positions on the DNA backbone in the non-contacted region. The positions of these "phosphate pockets" create requires that the DNA conform to the B' state in order to properly align with P22R NTD. The "phosphate pockets" are created by the δNH_2 of Asn49 and the backbone nitrogen of Asn46 on both the L and R monomers which chelate the O2 of the phosphate of residue 13 and 33 of the DNA, respectively (Figure 4.6). Finally, the N-terminal of helix four is located directly over the minor groove of the non-contacted region. The position of helix four directs the positive dipole of the helix toward the minor groove of the non-contacted region (Figure 4.8). The combination of the "phosphate pocket" and the alignment of helix four create an electrostatic surface that is complimentary to the B' state of the DNA (Figure 4.9).

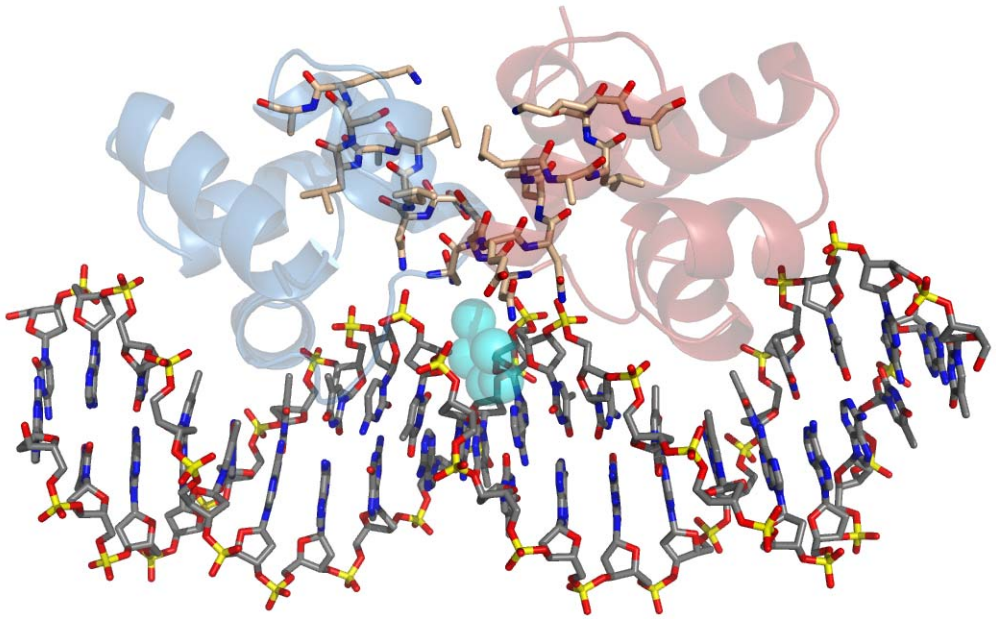

Figure 4.8. Helix four interaction with the induced B'-state minor groove. Helix 4 of the protein and the DNA are shown in sticks. The L monomer is shown in blue and the R monomer is shown in red ribbons. Waters of the spine of hydration are shown in cyan spheres. The water and protein ribbons are set partially transparent for clarity.

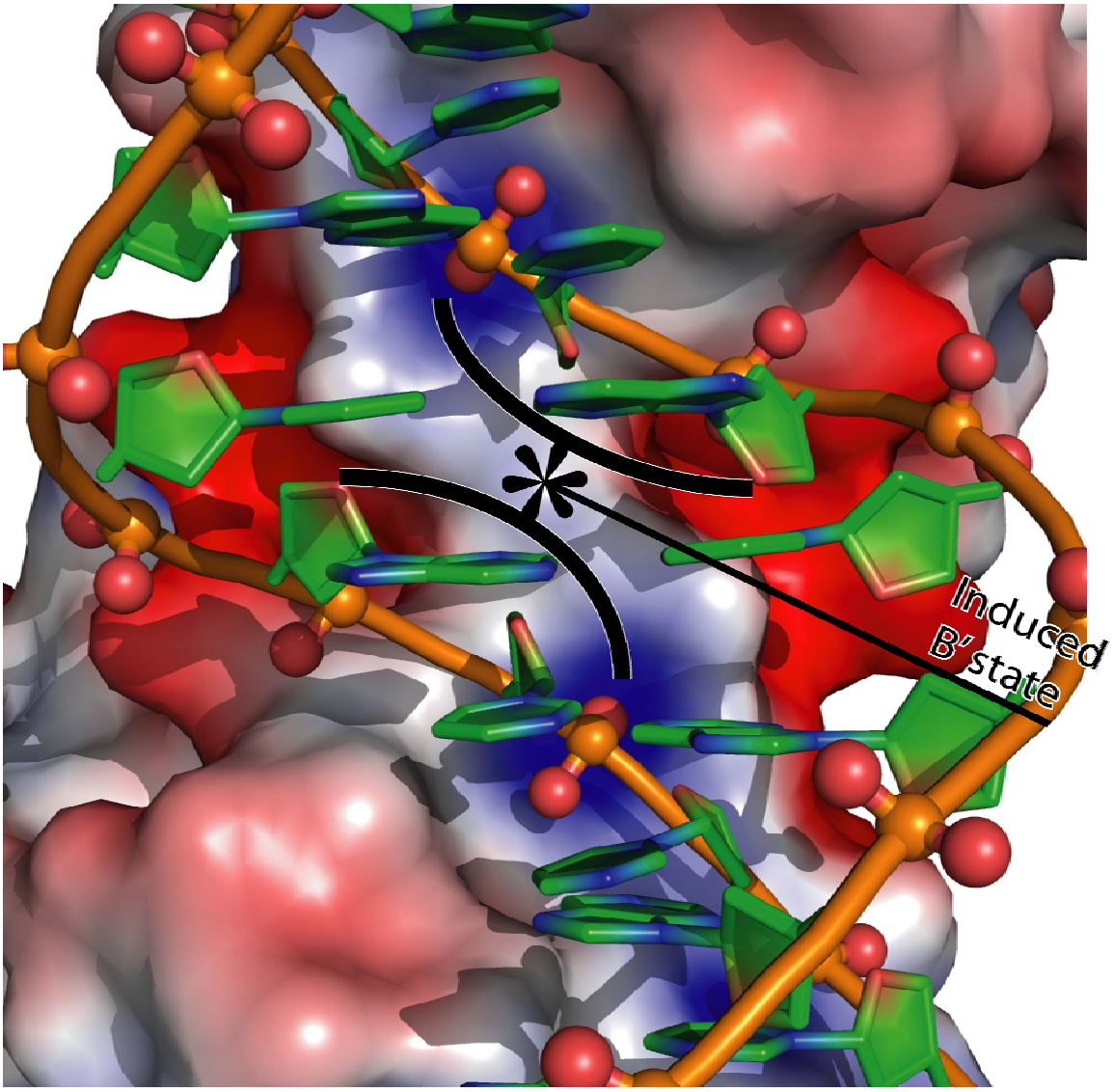

Figure 4.9. Graphical representation of ESP calculated at the solvent accessible surface area of the P22R NTD alone. P22R NTD shown as surface colored red, -3 kT/e^-, to blue +3 kT/e^-. DNA backbone shown in tubes with phosphate oxygens shown in red spheres. DNA bases shown as cartoon. Black lines indicate the narrowing of the minor groove in the non-contacted region.

Positions of monovalent cations in A-tracts.

A total of ten sites are indicated as positions of localized monovalent cations in the P22R NTD-DNA^{9T}/Tl^{+} complex and two sites are identified as sites of localized monovalent cations in the P22R NTD-DNA^{9C}/Rb^{+} derivative. As expected from previous work, cation localization sites in the major groove when associated with CG base pairs and in the minor groove when associated with AT base pairs. In particular, the DNA operator described here contains two crystallographically independent A-tracts. In each, the primary hydration layer, which is the base of the spine of hydration is seen to be a hybrid of cation and water, each partially occupied. This model of DNA-cation interactions, first proposed in 1998 (*38*) and contested by Dickerson (*130*), now seems well-established.

The identification of localized monovalent cation sites indicates the experimental conditions were sufficient to identify the positions of cations (*38, 53, 54, 56, 84*). The positions of the ten monovalent cations in the P22R NTD-DNA^{9T}/Tl^{+} are equivalent to previously identified cation sites of other x-ray crystallography and NMR studies. The highest occupied sites occur at the terminal ends of the operator and are coordinated in the major groove by the N7 and O6 of guanine. The approximately equivalent site was observed at a GpC step in a Tl^{+} derivative of the Dickerson-Drew dodecamer, and was shown to be the highest occupied monovalent cation site within the structure (*54*). The coordination of the Tl^{+MG2} and the Tl$^{+MG2'}$ sites in the P22R NTD-DNA^{9T}/Tl^{+} is enhanced due to the nick in the backbone at the junction of the symmetry related operators (see Results). The nick in the backbone allows the over twisting of the base pairs so that two cross-strand guanines can coordinate cation. The final site identified by

the anomalous signal in the P22R NTD-DNA^{9T}/Tl^+ complex is also in the major groove and is coordinated by the O6 and N7 of a guanine at the position 7. At this position in the operator, the minor groove widens and the base pairs roll toward the major groove.

Cation Localization and the Minor Groove of the Non-contacted Region

The structures of the P22R NTD-DNA^{9T}/Tl^+ and the P22R NTD-DNA^{9C}/Rb^+ derivatives here, give no indication of localized monovalent cations in the minor groove of the non-contacted region.

The other seven positions of cation localization were not identified by anomalous scattering, but by the isomorphous difference between a native P22R NTD-DNA^{9T}/Na^+ derivative (chapter 2) and the P22R NTD-DNA^{9T}/Tl^+ derivative. These additional seven sites identified by the F_{Tl+}-F_{Na+} maps are located in the minor groove of the A-tracts at the end of the operators. The equivalent positions of the cations localized in the minor groove of the A-tracts on each end of the operator have been previously identified by crystallography and/or NMR experiments in the A-tracts of unbound DNA, and are associated with the narrow minor groove of the B' state.

Only two cation localization sites were determined from the P22R NTD-DNA^{9C}/Rb^+ derivatives. These two sites correspond to the two highest sites at the terminal base pairs of the P22R NTD-DNA^{9T}/Tl^+ derivatives. It is likely that only two sites are observed in the Rb^+ derivative due to the weaker anomalous signal of Rb^+ compared to that of Tl^+, the lower quality data of the Rb^+ derivative, and the reduced resolution of the Rb^+ derivative. Furthermore, an isomorphous native derivative was not available for the DNA^{9C} complex.

Calculations of the ESP of the P22R NTD-DNA complex indicate that repressor binding increases the negative potential of the minor groove in the non-contacted region. The potential in the minor groove of the non-contacted region is more negative than that of the A-tracts at each end of the DNA. Cations are observed in the minor groove of the A-tracts on each end of the operator and not in the minor groove of the non-contacted region. However, the inability to identify a cation in the non-contacted region does not eliminate the possibility that a cation may be present. NMR data has shown that Na^+ binds with a much greater preference than Rb^+ in the minor groove of A-tracts (*84*). When Na^+ is present in the crystallization conditions of the P22R NTD-DNA^{9T}/Tl^+, no anomalous or isomorphous peaks are observed (data not shown). Great links were taken to insure that no Na^+ was present in the crystallization of the P22R NTD-DNA complexes. However, the ubiquitous nature of Na^+ makes it impossible to rule out its presence.

NMR experiments have shown that NH_4^+ is similar way to Na^+ in the minor groove interactions of A-tracts. The DNA oligomers used in the crystallization were obtained as NH_4^+ salts, thus a small amount of NH_4^+ was present in the crystallization conditions. Therefore, if the cation coordination in the minor groove of non-contacted region is highly selective for Na^+, the K^+ mimics may have been displaced.

The coordination in the minor groove of the non-contacted region differs from that in the minor groove of the A-tracts. The cations observed in the minor groove of A-tracts are coordinated by one or more oxygens. No oxygen atoms are present in the minor groove of the non-contacted region of the operators studied thus far. The lack of an oxygen atom to coordinate a cation in the minor groove may result in a delocalization

of the cations. The delocalization of the cations in the minor groove would result in an even distribution throughout the groove, and would not be detectible with x-ray crystallography.

APPENDIX A

Local Base Pair Parameters of the P22R NTD-DNA^{9T}.*

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
-1R C-G	0.00	-0.40	-0.10	4.20	-10.73	1.59
1R A-T	-0.15	-0.13	0.02	-0.27	-12.60	1.61
2R T-A	-0.09	-0.13	0.05	2.98	-10.72	0.66
3R T-A	-0.03	-0.11	0.06	2.52	-13.62	3.18
4R T-A	-0.02	-0.16	-0.06	-1.45	-12.80	1.49
5R A-T	0.04	-0.12	-0.15	-5.83	-2.60	0.50
6R A-T	-0.08	-0.15	0.12	4.07	-12.75	6.18
7R G-C	-0.25	-0.18	0.04	0.65	-15.98	1.67
8R A-T	0.10	-0.06	0.08	2.69	-6.41	0.33
9R T-A	-0.12	-0.10	-0.05	6.60	-8.92	1.95
9L A-T	-0.05	-0.17	0.11	-7.39	-4.22	2.53
8L T-A	-0.07	-0.09	0.20	-2.28	-10.08	-1.03
7L C-G	0.14	-0.17	0.16	-4.65	-17.52	0.94
6L T-A	-0.03	-0.15	-0.05	-1.20	-9.67	2.84
5L T-A	0.04	-0.17	0.00	4.88	-2.90	-0.62
4L A-T	0.02	-0.14	0.01	4.21	-13.89	0.94
3L A-T	-0.11	0.01	0.02	-0.59	-13.23	3.49
2L A-T	0.11	-0.02	0.11	-0.69	-13.55	0.20
1L T-A	0.34	-0.12	-0.09	-0.03	-9.71	0.08
-1L A-T	-0.99	-0.79	0.17	2.91	-8.32	-6.22
ave.	-0.06	-0.17	0.03	0.57	-10.51	1.12
s.d.	0.25	0.17	0.10	3.73	4.08	2.37

* Calculated with 3DNA.

APPENDIX B

Local Base Pair Step Parameters of the P22R NTD-DNA^{9T}.*

Step	Shift	Slide	Rise	Tilt	Roll	Twist
-1R to 1R CA/TG	-0.06	0.13	3.41	1.70	8.52	40.36
1R to 2R AT/AT	0.20	-0.72	3.24	0.02	-1.69	31.35
2R to 3R TT/AA	-0.07	-0.44	3.25	0.10	-2.69	36.96
3R to 4R TT/AA	-0.09	-0.44	3.29	0.35	-1.40	38.11
4R to 5R TA/TA	0.17	-0.22	3.43	-0.54	0.76	37.57
5R to 6R AA/TT	0.43	0.16	3.12	-3.12	10.58	30.80
6R to 7R AG/CT	0.47	-0.64	3.37	-0.61	6.58	32.27
7R to 8R GA/TC	-0.97	-0.10	3.18	-1.24	-0.61	40.46
8R to 9R AT/AT	0.24	-0.43	3.27	2.04	-4.64	33.33
9R to 9L TA/TA	0.27	-0.06	3.65	-0.78	-4.58	45.87
9L to 8L AT/AT	-0.74	-0.66	3.23	-2.26	-0.11	30.03
8L to 7L TC/GA	0.75	-0.13	3.26	2.09	-1.48	39.42
7L to 6L CT/AG	-0.57	-0.23	3.26	1.88	8.02	31.89
6L to 5L TT/AA	-0.27	0.22	3.25	-0.13	8.27	32.29
5L to 4L TA/TA	-0.14	-0.12	3.34	1.49	3.00	35.88
4L to 3L AA/TT	0.27	-0.38	3.33	0.78	-0.88	36.73
3L to 2L AA/TT	0.10	-0.50	3.24	-1.35	-3.77	38.97
2L to 1L AT/AT	-0.12	-0.93	3.30	0.49	0.53	31.56
1L to -1L TA/TA	0.20	-0.33	3.19	-5.01	8.22	35.11
ave.	0.00	-0.31	3.29	-0.22	1.72	35.73
s.d.	0.42	0.31	0.11	1.85	5.02	4.22

* Calculated with 3DNA.

APPENDIX C

Local Base Pair Helical Parameters of P22R NTD-DNA^{9T} .*

Step	X-disp	Y-disp	H-Rise	Incl.	Tip	H-Twist
-1R to 1R CA/TG	-0.78	0.28	3.37	12.17	-2.42	41.24
1R to 2R AT/AT	-1.02	-0.37	3.27	-3.13	-0.03	31.39
2R to 3R TT/AA	-0.34	0.13	3.27	-4.23	-0.15	37.06
3R to 4R TT/AA	-0.50	0.19	3.30	-2.14	-0.54	38.13
4R to 5R TA/TA	-0.44	-0.33	3.42	1.18	0.83	37.58
5R to 6R AA/TT	-1.48	-1.28	2.96	19.16	5.65	32.67
6R to 7R AG/CT	-2.26	-0.94	3.17	11.68	1.09	32.92
7R to 8R GA/TC	-0.08	1.26	3.21	-0.87	1.79	40.49
8R to 9R AT/AT	0.03	-0.07	3.30	-8.03	-3.53	33.70
9R to 9L TA/TA	0.36	-0.41	3.63	-5.86	1.00	46.09
9L to 8L AT/AT	-1.25	0.97	3.28	-0.21	4.36	30.11
8L to 7L TC/GA	-0.02	-0.87	3.29	-2.19	-3.09	39.50
7L to 6L CT/AG	-1.76	1.32	3.08	14.31	-3.36	32.91
6L to 5L TT/AA	-0.99	0.45	3.21	14.58	0.22	33.30
5L to 4L TA/TA	-0.63	0.43	3.31	4.85	-2.41	36.03
4L to 3L AA/TT	-0.48	-0.33	3.34	-1.40	-1.24	36.75
3L to 2L AA/TT	-0.29	-0.31	3.27	-5.64	2.02	39.17
2L to 1L AT/AT	-1.80	0.32	3.28	0.97	-0.91	31.56
1L to -1L TA/TA	-1.63	-0.99	2.99	13.33	8.11	36.36
	-0.81	-0.03	3.26	3.08	0.39	36.16
	0.73	0.74	0.15	8.38	3.09	4.07

* Calculated with 3DNA.

APPENDIX D

Bridging Water Molecules (Protein-W-DNA) in P22R NTD-DNA^{9T}

water #	DNA/protein atom	Distance (Å)
1	T32 O2P	2.7
1	E44R CO	3.2
1	P45R CO	2.9
1	G47L N	3.4
2	C33 P01	2.8
2	Q37R OE	2.7
2	T43R OG	2.7
3	T12 O2P	2.7
3	P45L CO	2.8
3	E44L CO	3.3
3	G47R N	3.4
4	Q21R OE	2.7
4	N32R ND	2.8
6	A6 N7	2.9
6	A6 N6	3.1
6	Q37R NE	3.3
12	C33 N4	3
12	E42R OE	2.8
14	T5 PO1	2.8
14	E42R OE	3.5
14	R40R NH2	3.6
18	A36 N6	3
18	A36 N7	3
18	V33R CG	3.6
19	C13 PO1	2.7
19	Q37L OE	2.7
19	T43L OG	2.7
23	T3 PO1	2.7
23	A22R N	3
23	Q21R N	3.4
24	T4 O4	2.7
24	A37 N6	3.5
24	N32R ND	2.9
24	V33R CG	3.9
30	T5 C5M	3.1
30	R40R NH2	2.9
30	E42R OE	2.5
30	E42R OE	3
31	T3 O2	2.7
31	K15R NZ	2.8
38	T32 PO2	3
38	T32 O3'	3.1
38	N46L ND	2.9
44	A26 N7	2.9
44	A26 N6	3.2
44	E42LOE	3
54	C13 N4	3

54	E42L OE	2.6
68	T23 PO1	2.6
68	A22L N	2.9
68	Q21L N	3.3
71	T32 PO1	3.3
71	E42R CO	2.7
77	T35 PO1	2.6
77	S31R CB	3.6
96	T25 PO1	2.7
96	R40L NH	2.9
105	T34 PO2	2.6
105	S31R N	3.7
106	T24 04	2.7
106	A17 N6	3.4
106	N32L ND	3.1
106	V33L CG	3.8
107	T4 O2	2.7
107	R11R NH2	3.6
112	T15 PO1	2.6
112	S31L CB	3.3
116	T14 PO2	2.7
116	G29L CO	3.3
119	A11 O3'	3.1
119	T12 PO2	3.2
119	N46R ND	3.1
123	A16 N7	3
123	A16 N6	3.2
123	V33L CG	3.8
130	T4 O2	2.7
130	R11R NH2	3.2
136	T32 PO2	2.7
136	E44R OE	3.8
137	T5 PO1	2.86
137	R40R NH2	2.9
139	A27 N6	3.1
139	A27 N7	3.4
139	E42L OE	3.7
182	T12 PO2	2.7
182	E44L OE	3.6
212	C33 O5'	2.9
212	T32 O3'	3.2
212	C33 O4'	3.4
212	N46R ND	3.6
226	T34 PO2	2.6
226	G29R CO	2.6
268	T25 PO1	3.7
268	E42L OE	2.8

APPENDIX E

Contacts of the Valine 33 Sidechain with DNA in P22R NTD-DNA^{9T}

Val33 atom	DNA base atom	Distance Å
CG1	5L (T) O4	3.27
CG1	5R (T) O4	3.40
CG1	4L (T) O4	3.31
CG1	4R (T) O4	3.58
CG1	6L (T) C5M	4.06
CG1	6R (T) C5M	3.98
CG1	4L (T) C5M	4.06
CG1	4R (T) C5M	4.18
CG2	5L (T) O4	3.45
CG2	5R (T) O4	3.54
CG2	4L (A) N6	3.55
CG2	4R (A) N6	3.54
CG2	5L (T) C5M	3.66
CG2	5R (T) C5M	3.89
CG2	3L (T) C5M	4.35
CG2	3R (T) C5M	4.28
CB	5L (T) C5M	3.84
CB	5R (T) C5M	4.01
CB	6L (T) C5M	4.18
CB	6R (T) C5M	4.22
CA	3L (T) C5M	4.26
CA	3R (T) C5M	4.32

REFERENCES

1. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) Sequence-Specific Recognition of Double Helical Nucleic Acids by Proteins, *Proc. Natl. Acad. Sci. U.S.A. 73*, 804-808.

2. Von Hippel, P. H., and Berg, O. G. (1986) On the Specificity of DNA-Protein Interactions, *Proceedings of the National Academy of Sciences of the United States of America 83*, 1608-1612.

3. Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z., and Crothers, D. M. (2004) Predicting indirect readout effects in protein-DNA interactions, *Proc. Natl. Acad. Sci. U. S. A. 101*, 8337-8341. Epub 2004 May 8317.

4. Anderson, J. E., Ptashne, M., and Harrison, S. C. (1987) Structure of the Repressor Operator Complex of Bacteriophage 434, *Nature 326*, 846-852.

5. Koudelka, G. B., Harrison, S. C., and Ptashne, M. (1987) Effect of Non-Contacted Bases on the Affinity of 434 Operator for 434 Repressor and Cro, *Nature 326*, 886-888.

6. Mauro, S. A., and Koudelka, G. B. (2004) Monovalent cations regulate DNA sequence recognition by 434 repressor, *J Mol Biol. 340*, 445-457.

7. DeDecker, B. S., Obrien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P., and Sigler, P. B. (1996) The crystal structure of a hyperthermophilic archaeal TATA-box binding protein, *Journal of Molecular Biology 264*, 1072-1084.

8. Kim, Y. C., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal-Structure of a Yeast Tbp Tata-Box Complex, *Nature 365*, 512-520.

9. Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., and Burley, S. K. (1996) Crystal structure of a human TATA box-binding protein/TATA element complex, *Proc. Natl. Acad. Sci. U.S.A. 93*, 4862-4867.

10. Faiger, H., Ivanchenko, M., and Haran, T. E. (2007) Nearest-neighbor non-additivity versus long-range non-additivity in TATA-box structure and its

implications for TBP-binding mechanism, *Nucleic Acids Res. 35*, 4409-4419. Epub 2007 Jun 4418.

11. Bareket-Samish, A., Cohen, I., and Haran, T. E. (1998) Direct versus indirect readout in the interaction of the trp repressor with non-canonical binding sites, *Journal of Molecular Biology 277*, 1071-1080.

12. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature 335*, 321-329.

13. Somers, W. S., and Phillips, S. E. (1992) Crystal structure of the met repressor-operator complex at 2.8 A resolution reveals DNA recognition by beta-strands, *Nature. 359*, 387-393.

14. Mouw, K. W., and Rice, P. A. (2007) Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending protein Hbb, *Molecular Microbiology 63*, 1319-1330.

15. Yang, C. C., and Nash, H. A. (1989) The Interaction of Escherichia-Coli Ihf Protein with Its Specific Binding-Sites, *Cell 57*, 869-880.

16. Gabrielsen, O. S., Sentenac, A., and Fromageot, P. (1991) Specific DNA binding by c-Myb: evidence for a double helix-turn-helix-related motif, *Science. 253*, 1140-1143.

17. Dangi, B., Pelupessey, P., Martin, R. G., Rosner, J. L., Louis, J. M., and Gronenborn, A. M. (2001) Structure and dynamics of MarA-DNA complexes: an NMR investigation, *J. Mol. Biol. 314*, 113-127.

18. Blakaj, D. M., Kattamuri, C., Khrapunov, S., Hegde, R. S., and Brenowitz, M. (2006) Indirect readout of DNA sequence by papillomavirus E2 proteins depends upon net cation uptake, *Journal of Molecular Biology 358*, 224-240.

19. Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) Crystal-Structure at 1.7-Angstrom of the Bovine Papillomavirus-1 E2 DNA-Binding Domain Bound to Its DNA Target, *Nature 359*, 505-512.

20. Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements, *Cell. 75*, 567-578.

21. Chen, S., Vojtechovsky, J., Parkinson, G. N., Ebright, R. H., and Berman, H. M. (2001) Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking, *J. Mol. Biol. 314*, 63-74.

22. Joshi, H. K., Etzkorn, C., Chatwell, L., Bitinaite, J., and Horton, N. C. (2006) Alteration of sequence specificity of the type II restriction endonuclease HincII through an indirect readout mechanism, *J. Biol. Chem. 281*, 23852-23869.

23. Wu, L., Vertino, A., and Koudelka, G. B. (1992) Non-contacted bases affect the affinity of synthetic P22 operators for P22 repressor, *J. Biol. Chem. 267*, 9134-9139.

24. Lamoureux, J. S., Maynes, J. T., and Glover, J. N. (2004) Recognition of 5'-YpG-3' sequences by coupled stacking/hydrogen bonding interactions with amino acid residues, *J. Mol. Biol. 335*, 399-408.

25. Schumann, W., Lindenblatt, E., and Bade, E. G. (1976) Bacteriophage-specific DNA-binding proteins in P22-lysogenic and in P22-infected Salmonella typhimurium, *J Virol. 20*, 334-338.

26. Poteete, A. R., and Ptashne, M. (1982) Control of Transcription by the Bacteriophage P22 Repressor, *Journal of Molecular Biology 157*, 21-48.

27. Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A., and Maurer, R. (1975) Recognition Sequences of Repressor and Polymerase in Operators of Bacteriophage-Lambda, *Cell 5*, 109-113.

28. Wu, L., and Koudelka, G. B. (1993) Sequence-dependent differences in DNA structure influence the affinity of P22 operator for P22 repressor, *J Biol Chem. 268*, 18975-18981.

29. Hilchey, S. P., and Koudelka, G. B. (1997) DNA-based loss of specificity mutations. Effects of DNA sequence on the contacted and non-contacted base preferences of bacteriophage P22 repressor, *J Biol Chem. 272*, 1646-1653.

30. Wharton, R. P., and Ptashne, M. (1985) Changing the binding specificity of a repressor by redesigning an alpha-helix, *Nature. 316*, 601-605.

31. Sevilla-Sierra, P., Otting, G., and Wuthrich, K. (1994) Determination of the nuclear magnetic resonance structure of the DNA-binding domain of the P22 c2 repressor (1 to 76) in solution and comparison with the DNA-binding domain of the 434 repressor, *J. Mol. Biol. 235*, 1003-1020.

32. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) Recongition of a DNA Operator by the Repressor of Phage 434: A View at High Resolution, *Science 242*, 899-907.

33. Shimon, L. J., and Harrison, S. C. (1993) The phage 434 OR2/R1-69 complex at 2.5 A resolution, *J Mol Biol. 232*, 826-838.

34. Rodgers, D. W., and Harrison, S. C. (1993) The complex between phage 434 repressor DNA-binding domain and operator site OR3: structural differences between consensus and non-consensus half-sites, *Structure. 1*, 227-240.

35. Hilchey, S. P., Wu, L., and Koudelka, G. B. (1997) Recognition of nonconserved bases in the P22 operator by P22 repressor requires specific interactions between repressor and conserved bases, *J. Biol. Chem. 272*, 19898-19905.

36. Hud, N. V., and Plavec, J. (2003) A unified model for the origin of DNA sequence-directed curvature, *Biopolymers 69*, 144-158.

37. Drew, H. R., and Dickerson, R. E. (1981) Structure of a B-DNA dodecamer. III. Geometry of Hydration, *J. Mol. Biol. 151*, 535-556.

38. Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium, *Biochemistry 37*, 8341-8355.

39. Kopka, M. L., Fratini, A. V., Drew, H. R., and Dickerson, R. E. (1983) Ordered water structure around a B-DNA dodecamer. A quantitative study, *J. Mol. Biol. 163*, 129-146.

40. Woods, K., Lan, T., McLaughlin, L. W., and Williams, L. D. (2003) The Role of Minor Groove Functional Groups in DNA Hydration, *Nucleic Acids Res 31*, 1536-1540.

41. Lan, T., and McLaughlin, L. W. (2000) Minor groove hydration is critical to the stability of DNA duplexes, *J. Am. Chem. Soc. 122*, 6512-6513.

42. Lan, T., and McLaughlin, L. W. (2001) Minor groove functional groups are critical for the B-form conformation of duplex DNA, *Biochemistry 40*, 968-976.

43. Sun, Z. H., Chen, D. L., Lan, T., and McLaughlin, L. W. (2002) Importance of minor groove functional groups for the stability of DNA duplexes, *Biopolymers 65*, 211-217.

44. Watson, J. D., and Crick, F. H. C. (1953) Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid, *Nature 171*, 737-738.

45. Rich, A. (1993) DNA Comes in Many Forms, *Gene 135*, 99-109.

46. Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides, *Q. Rev. Biophys. 11*, 179-246.

47. Anderson, C. F., Record, M. T., Jr., and Hart, P. A. (1978) Sodium-23 NMR studies of cation-DNA interactions, *Biophys Chem 7*, 301-316.

48. Hamelberg, D., Williams, L. D., and Wilson, W. D. (2001) Influence of the Dynamic Positions of Cations on the Structure of the DNA Minor Groove: Sequence-Dependent Effects, *J. Am. Chem. Soc. 123*, 7745-7755.

49. Williams, L. D., and Maher, L. J. (2000) Electrostatic Mechanisms of DNA Deformation, *Annu. Rev. Biophys. Biomol. Struct. 29*, 497-521.

50. Hud, N. V., and Polak, M. (2001) DNA-cation interactions: The major and minor grooves are flexible ionophores, *Curr. Opin. Struct. Biol. 11*, 293-301.

51. Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D., and Williams, L. D. (2000) Cations mediate B-DNA conformational heterogeneity, *J. Am. Chem. Soc. 122*, 11048-11056.

52. Marincola, F. C., Denisov, V. P., and Halle, B. (2004) Competitive Na+ and Rb+ binding in the minor groove of DNA, *J. Am. Chem. Soc. 126*, 6739-6750.

53. Woods, K., McFail-Isom, L., Sines, C. C., Howerton, S. B., Stephens, R. K., and Williams, L. D. (2000) Monovalent Cations Sequester within the A-Tract Minor Groove of [d(CGCGAATTCGCG)]2, *J. Am. Chem. Soc. 122*, 1546-1547.

54. Howerton, S. B., Sines, C. C., VanDerveer, D., and Williams, L. D. (2001) Locating Monovalent Cations in the Grooves of B-DNA, *Biochemistry 40*, 10023-10031.

55. Tereshko, V., Minasov, G., and Egli, M. (1999) A "hydrat-ion" spine in a B-DNA minor groove, *J. Am. Chem. Soc. 121*, 3590-3595.

56. Hud, N. V., Sklenar, V., and Feigon, J. (1999) Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending, *J. Mol. Biol. 286*, 651-660.

57. Hud, N. V., Schultze, P., and Feigon, J. (1998) Ammonium ion as an NMR probe for monovalent cation coordination sites of DNA quadruplexes, *J. Am. Chem. Soc. 120*, 6403-6404.

58. Shui, X., Sines, C., McFail-Isom, L., VanDerveer, D., and Williams, L. D. (1998) Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations, *Biochemistry 37*, 16877-16887.

59. Lan, T., and McLaughlin, L. W. (2001) Minor groove functional groups are critical for the B-form conformation of duplex DNA, *Biochemistry 40*, 968-976.

60. Ptashne, M. (2005) Regulation of transcription: from lambda to eukaryotes, *Trends in Biochemical Sciences 30*, 275-279.

61. Kaiser, A. D. (1957) Mutations in a Temperate Bacteriophage Affecting Its Ability to Lysogenize Escherichia-Coli, *Virology 3*, 42-61.

62. Poteete, A. R., Ptashne, M., Ballivet, M., and Eisen, H. (1980) Operator Sequences of Bacteriophage-P22 and Bacteriophage-21, *Journal of Molecular Biology 137*, 81-91.

63. Haran, T. E., and Crothers, D. M. (1989) Cooperativity in A-tract structure and bending properties of composite TnAn blocks, *Biochemistry. 28*, 2763-2767.

64. De Anda, J., Poteete, A. R., and Sauer, R. T. (1983) P22 c2 repressor. Domain structure and function, *J. Biol. Chem. 258*, 10536-10542.

65. Amann, E., Brosius, J., and Ptashne, M. (1983) Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in Escherichia coli, *Gene 25*, 167-178.

66. Coulondre, C., and Miller, J. H. (1977) Genetic studies of the lac repressor. IV. Mutagenic specificity in the lacI gene of Escherichia coli, *J. Mol. Biol. 117*, 577-606.

67. Otwinowski, Z. (1993) Data Collection and Processing, in *Data Collection and Processing*, pp 56-62, Science and Engineering Research Council, Warrington, United Kingdom.

68. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, in *Methods in Enzymol., Macromolecular Crystallography* (Carter, J., C.W., and Sweet, R. M., Eds.), pp 307-326, Academic Press, New York.

69. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. Sect. D-Biol. Crystallogr. 54*, 905-921.

70. Kleywegt, G. J., and Jones, T. A. (1999) Software for handling macromolecular envelopes, *Acta Crystallogr. Sect. D-Biol. Crystallogr. 55*, 941-944.

71. Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins, *Adv Protein Chem 23*, 283-438.

72. Mack, D. R., Chiu, T. K., and Dickerson, R. E. (2001) Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: A comparative analysis of T-A and A-T steps within A-tracts, *Journal of Molecular Biology 312*, 1037-1049.

73. Williams, L. D. (2005) Between Objectivity and Whim: Nucleic Acid Structural Biology, in *Topics in Current Chemistry* (Chaires, J. B., and Waring, M., Eds.), pp 77-88, Springer, Heidelberg.

74. McFail-Isom, L., Sines, C., and Williams, L. D. (1999) DNA Structure: Cations in Charge?, *Current Op. Struct. Biol. 9*, 298-304.

75. Poteete, A. R., and Ptashne, M. (1982) Control of transcription by the bacteriophage P22 repressor, *J. Mol. Biol. 157*, 21-48.

76. Arnott, S., and Hukins, D. W. L. (1972) Optimized Parameters for a-DNA and B-DNA, *Biochemical and Biophysical Research Communications 47*, 1504.

77. Arnott, S., and Hukins, D. W. L. (1972) Optimized Parameters for a-DNA and B-DNA, *Biochemical and Biophysical Research Communications 47*, 1504-&.

78. Wu, H.-M., and Crothers, D. M. (1984) The locus of sequence-directed and protein-induced DNA bending, *Nature 308*, 509-513.

79. Marini, J. C., Levene, S. D., Crothers, D. M., and Englund, P. T. (1982) Bent helical structure in kinetoplast DNA, *Proc. Natl. Acad. Sci., U.S.A. 79*, 7664-7668.

80. Zinkel, S. S., and Crothers, D. M. (1987) DNA bend direction by phase sensitive detection, *Nature 328*, 178-181.

81. Burkhoff, A. M., and Tullius, T. D. (1987) The unusual conformation adopted by the adenine tracts in kinetoplast DNA, *Cell 48*, 935-943.

82. Alexeev, D. G., Lipanov, A. A., and Skuratovskii, I. Y. (1987) Poly(dA)-poly(dT) is a B-type double helix with a distinctively narrow minor groove, *Nature 325*, 821-823.

83. Woods, K. K., Maehigashi, T., Howerton, S. B., Sines, C. C., Tannenbaum, S., and Williams, L. D. (2004) High-resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]2, *J. Am. Chem. Soc. 126*, 15330-15331.

84. Cesare Marincola, F., Denisov, V. P., and Halle, B. (2004) Competitive Na(+) and Rb(+) binding in the minor groove of DNA, *J Am Chem Soc 126*, 6739-6750.

85. Wing, R., Drew, H., Takano, T., Broka, C., Takana, S., Itakura, K., and Dickerson, R. E. (1980) Crystal structure analysis of a complete turn of B-DNA, *Nature 287*, 755-758.

86. Dickerson, R. E., and Drew, H. R. (1981) Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure, *J. Mol. Biol. 149*, 761-786.

87. Rhodes, D. (1979) Nucleosome Cores Reconstituted from Poly (Da-Dt) and the Octamer of Histones, *Nucleic Acids Research 6*, 1805-1816.

88. Simpson, R. T., and Kunzler, P. (1979) Chromatin and Core Particles Formed from the Inner Histones and Synthetic Polydeoxyribonucleotides of Defined Sequence, *Nucleic Acids Research 6*, 1387-1415.

89. Herrera, J. E., and Chaires, J. B. (1989) A Premelting Conformational Transition in Poly(dA)-Poly(dT) Coupled to Daunomycin Binding, *Biochemistry 28*, 1993-2000.

90. Chan, S. S., Breslauer, K. J., Austin, R. H., and Hogan, M. E. (1993) Thermodynamics and premelting conformational changes of phased (dA)5 tracts, *Biochemistry 32*, 11776-11784.

91. Augustyn, K. E., Wojtuszewski, K., Hawkins, M. E., Knutson, J. R., and Mukerji, I. (2006) Examination of the premelting transition of DNA A-tracts using a fluorescent adenosine analogue, *Biochemistry 45*, 5039-5047.

92. Nelson, H. C. M., Finch, J. T., Luisi, B. F., and Klug, A. (1987) The Structure of an Oligo(dA)•Oligo(dT) Tract and its Biological Implications, *Nature 330*, 221-226.

93. Coll, M., Frederick, C. A., Wang, A. H.-J., and Rich, A. (1987) A Bifurcated Hydrogen-Bonded Conformation in the d(A•T) Base Pairs of the DNA

Dodecamer d(CGCAAATTTGCG) and its Complex with Distamycin, *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385-8389.

94. Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Cascio, D., Gunsalus, R. P., and Dickerson, R. E. (1998) NarL dimerization? Suggestive evidence from a new crystal form, *Biochemistry 37*, 3665-3676.

95. Li, T., Jin, Y. S., Vershon, A. K., and Wolberger, C. (1998) Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer in complex with DNA containing an A-tract, *Nucleic Acids Research 26*, 5707-5718.

96. Ke, A. L., Mathias, J. R., Vershon, A. K., and Wolberger, C. (2002) Structural and thermodynamic characterization of the DNA binding properties of a triple alanine mutant of MAT alpha 2, *Structure 10*, 961-971.

97. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) Crystal-Structure of the Mata1/Mat-Alpha-2 Homeodomain Heterodimer Bound to DNA, *Science 270*, 262-269.

98. Santelli, E., and Richmond, T. J. (2000) Crystal structure of MEF2A core bound to DNA at 1.5 angstrom resolution, *Journal of Molecular Biology 297*, 437-449.

99. Mondragon, A., and Harrison, S. C. (1991) The phage 434 Cro/OR1 complex at 2.5 A resolution, *J Mol Biol. 219*, 321-334.

100. Beamer, L. J., and Pabo, C. O. (1992) Refined 1.8 A crystal structure of the lambda repressor-operator complex, *J Mol Biol. 227*, 177-196.

101. Albright, R. A., and Matthews, B. W. (1998) Crystal structure of lambda-Cro bound to a consensus operator at 3.0 A resolution, *J. Mol. Biol. 280*, 137-151.

102. Jain, D., Kim, Y., Maxwell, K. L., Beasley, S., Zhang, R. G., Gussin, G. N., Edwards, A. M., and Darst, S. A. (2005) Crystal structure of bacteriophage lambda cII and its DNA complex, *Molecular Cell 19*, 259-269.

103. Chen, C. S. Y., White, A., Love, J., Murphy, J. R., and Ringe, D. (2000) Methyl groups of thymine bases are important for nucleic acid recognition by DtxR, *Biochemistry 39*, 10397-10407.

104. Beamer, L. J., and Pabo, C. O. (1992) Refined 1.8 A crystal structure of the lambda repressor-operator complex, *J. Mol. Biol. 227*, 177-196.

105. Wisedchaisri, G., Holmes, R. K., and Hol, W. G. J. (2004) Crystal structure of an IdeR-DNA complex reveals a conformational change in activated IdeR for base-specific interactions, *Journal of Molecular Biology 342*, 1155-1169.

106. Wisedchaisri, G., Chou, C. J., Wu, M. T., Roach, C., Rice, A. E., Holmes, R. K., Beeson, C., and Hol, W. G. J. (2007) Crystal structures, metal activation, and DNA-binding properties of two-domain IdeR from Mycobacterium tuberculosis, *Biochemistry 46*, 436-447.

107. Maris, A. E., Kaczor-Grzeskowiak, M., Ma, Z. C., Kopka, M. L., Gunsalus, R. P., and Dickerson, R. E. (2005) Primary and secondary modes of DNA recognition by the NarL two-component response regulator, *Biochemistry 44*, 14538-14552.

108. Blanco, A. G., Sola, M., Gomis-Ruth, F. X., and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator, *Structure 10*, 701-713.

109. Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., and Morikawa, K. (1996) Structure of a replication-terminator protein complexed with DNA, *Nature 383*, 598-603.

110. Oobatake, M., Kono, H., Wang, Y. F., and Sarai, A. (2003) Anatomy of specific interactions between lambda repressor and operator DNA, *Proteins-Structure Function and Genetics 53*, 33-43.

111. (1994) The CCP4 Suite: Programs for Protein Crystallography, *Acta Crystallogr. Sect. D-Biol. Crystallogr. D50*, 760-763.

112. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallographica Section D-Biological Crystallography 53*, 240-255.

113. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D-Biological Crystallography 60*, 2126-2132.

114. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures, *Journal of Applied Crystallography 26*, 283-291.

115. Watkins, D., Hsiao, C. L., Woods, K. K., Koudelka, G. B., and Williams, L. D. (2008) P22 c2 repressor-operator complex: Mechanisms of direct and indirect readout, *Biochemistry 47*, 2325-2338.

116. Watson, J. D., and Crick, F. H. (1974) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in Nature, number 4356 April 25, 1953, *Nature. 248*, 765.

117. Goodsell, D. S., Kaczor-Grzeskowiak, M., and Dickerson, R. E. (1994) The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps, *J. Mol. Biol. 239*, 79-96.

118. Arnott, S., and Hukins, D. W. L. (1972) Nucleotide Conformations, *Science 177*, 451-&.

119. Mendoza, R., Markovits, J., Jaffrezou, J. P., Muzard, G., and Le Pecq, J. B. (1990) DNase I Susceptibility of Bent DNA and Its Alteration by Ditercalinium and Distamycin, *Biochemistry 29*, 5035-5043.

120. Moulaei, T., Maehigashi, T., Lountos, G. T., Komeda, S., Watkins, D., Stone, M. P., Marky, L. A., Li, J. S., Gold, B., and Williams, L. D. (2005) Structure of B-DNA with cations tethered in the major groove, *Biochemistry 44*, 7458-7468.

121. Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferre-D'Amare, A. R., Strobel, S. A., and Doudna, J. A. (1998) A Specific Monovalent Metal Ion Integral to the AA Platform of the RNA Tetraloop Receptor, *Nat. Struct. Biol. 5*, 986-992.

122. Villeret, V., Huang, S., Fromm, H. J., and Lipscomb, W. N. (1995) Crystallographic Evidence for the Action of Potassium, Thallium, and Lithium ions on Fructose-1,6-bisphosphatase, *Proc. Natl. Acad. Sci. U.S.A. 92*, 8916-8920.

123. Wulfsberg, G. (1991) *Principles of Descriptive Inorganic Chemistry*, University Science Books, Sausalito, CA.

124. Morth, J. P., Pedersen, B. P., Toustrup-Jensen, M. S., Sorensen, T. L. M., Petersen, J., Andersen, J. P., Vilsen, B., and Nissen, P. (2007) Crystal structure of the sodium-potassium pump, *Nature 450*, 1043-U1046.

125. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome, *Proceedings of the National Academy of Sciences of the United States of America 98*, 10037-10041.

126. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations, *Nucleic Acids Research 32*, W665-W667.

127. Crick, F. H. C., and Magdoff, B. S. (1956) The theory of isomorphous replacement for protein crystals, *Acta Crystallogr. 9*, 901-908.

128. Howerton, S. B., Nagpal, A., and Williams, L. D. (2003) Surprising Roles for Electrostatic Interactions in DNA-Ligand Complexes, *Biopolymers 69*, 87-99.

129. Hud, N. V., Schultze, P., Sklenar, V., and Feigon, J. (1999) Binding Sites and Dynamics of Ammonium Ions in a Telomere Repeat DNA Quadruplex, *J. Mol. Biol. 285*, 233-243.

130. Chiu, T. K., Kaczor-Grzeskowiak, M., and Dickerson, R. E. (1999) Absence of Minor Groove Monovalent Cations in the Crosslinked Dodecamer CGCGAATTCGCG, *J. Mol. Biol. 292*, 589-608.